WIKIPEDIA

Artificial intelligence

Artificial intelligence (**AI**) is intelligence demonstrated by machines, as opposed to the **natural intelligence** displayed by humans or animals. Leading AI textbooks define the field as the study of "intelligent agents": any system that perceives its environment and takes actions that maximize its chance of achieving its goals.[a] Some popular accounts use the term "artificial intelligence" to describe machines that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving", however this definition is rejected by major AI researchers.[b]

AI applications include advanced web search engines (i.e. Google), recommendation systems (used by YouTube, Amazon and Netflix), understanding human speech (such as Siri or Alexa), self-driving cars (e.g. Tesla), and competing at the highest level in strategic game systems (such as chess and Go),[2] As machines become increasingly capable, tasks considered to require "intelligence" are often removed from the definition of AI, a phenomenon known as the AI effect.[3] For instance, optical character recognition is frequently excluded from things considered to be AI,[4] having become a routine technology.[5]

Artificial intelligence was founded as an academic discipline in 1956, and in the years since has experienced several waves of optimism,[6][7] followed by disappointment and the loss of funding (known as an "AI winter"),[8][9] followed by new approaches, success and renewed funding.[7][10] AI research has tried and discarded many different approaches during its lifetime, including simulating the brain, modeling human problem solving, formal logic, large databases of knowledge and imitating animal behavior. In the first decades of the 21st century, highly mathematical statistical machine learning has dominated the field, and this technique has proved highly successful, helping to solve many challenging problems throughout industry and academia.[11][10]

The various sub-fields of AI research are centered around particular goals and the use of particular tools. The traditional goals of AI research include reasoning, knowledge representation, planning, learning, natural language processing, perception and the ability to move and manipulate objects.[c] General intelligence (the ability to solve an arbitrary problem) is among the field's long-term goals.[12] To solve these problems, AI researchers use versions of search and mathematical optimization, formal logic, artificial neural networks, and methods based on statistics, probability and economics. AI also draws upon computer science, psychology, linguistics, philosophy, and many other fields.

The field was founded on the assumption that human intelligence "can be so precisely described that a machine can be made to simulate it".[d] This raises philosophical arguments about the mind and the ethics of creating artificial beings endowed with human-like intelligence. These issues have been explored by myth, fiction and philosophy since antiquity.[14] Science fiction and futurology have also suggested that, with its enormous potential and power, AI may become an existential risk to humanity.[15][16]

Contents

History

Precursors

Artificial beings with intelligence appeared as storytelling devices in antiquity,[17] and have been common in fiction, as in Mary Shelley's *Frankenstein* or Karel Čapek's *R.U.R.*[18] These characters and their fates raised many of the same issues now discussed in the ethics of artificial intelligence.[19]


Silver didrachma from Crete depicting Talos, an ancient mythical automaton with artificial intelligence

The study of mechanical or "formal" reasoning began with philosophers and mathematicians in antiquity. The study of mathematical logic led directly to Alan Turing's theory of computation, which suggested that a machine, by shuffling symbols as simple as "0" and "1", could simulate any conceivable act of mathematical deduction. This insight, that digital computers can simulate any process of formal reasoning, is known as the Church–Turing thesis.[20]

Cybernetics and brain simulation

The Church-Turing thesis, along with concurrent discoveries in neurobiology, information theory and cybernetics, led researchers to consider the possibility of building an electronic brain.[21] The first work that is now generally recognized as AI was McCullouch and Pitts' 1943 formal design for Turing-complete "artificial neurons".[22] By 1960, this approach was largely abandoned, although elements of it would be revived in the 1980s.

Symbolic AI

When access to digital computers became possible in the mid-1950s, AI research began to explore the possibility that human intelligence could be reduced to symbol manipulation. Approaches based on cybernetics or artificial neural networks were abandoned or pushed into the background.

The field of AI research was born at a workshop at Dartmouth College in 1956.[e][25] The attendees became the founders and leaders of AI research.[f] They and their students produced programs that the press described as "astonishing":[g] computers were learning checkers strategies, solving word problems in algebra, proving logical theorems and speaking English.[h][27] By the middle of the 1960s, research in the U.S. was heavily funded by the Department of Defense[28] and laboratories had been established around the world.[29]

Researchers in the 1960s and the 1970s were convinced that symbolic approaches would eventually succeed in creating a machine with artificial general intelligence and considered this the goal of their field.[30] Herbert Simon predicted, "machines will be capable, within twenty years, of doing any work a man can do".[31] Marvin Minsky agreed, writing, "within a generation ... the problem of creating 'artificial intelligence' will substantially be solved".[32]

They failed to recognize the difficulty of some of the remaining tasks. Progress slowed and in 1974, in response to the criticism of Sir James Lighthill[33] and ongoing pressure from the US Congress to fund more productive projects, both the U.S. and British governments cut off exploratory research in AI. The next few years would later be called an "AI winter", a period when obtaining funding for AI projects was difficult. [8]

In the early 1980s, AI research was revived by the commercial success of expert systems,[34] a form of AI program that simulated the knowledge and analytical skills of human experts. By 1985, the market for AI had reached over a billion dollars. At the same time, Japan's fifth generation computer project inspired the

U.S and British governments to restore funding for academic research.[7] However, beginning with the collapse of the Lisp Machine market in 1987, AI once again fell into disrepute, and a second, longer-lasting winter began.[9]

Early subsymbolic

Many researchers began to doubt that the symbolic approach would be able to imitate all the processes of human cognition, especially perception, robotics, learning and pattern recognition. A number of researchers began to look into "sub-symbolic" approaches to specific AI problems.[35] Robotics researchers, such as Rodney Brooks, rejected symbolic AI and focused on the basic engineering problems that would allow robots to move, survive, and learn their environment.[i] Interest in neural networks and "connectionism" was revived by Geoffrey Hinton, David Rumelhart and others in the middle of the 1980s.[40] Soft computing tools were developed in the 80s, such as neural networks, fuzzy systems, Grey system theory, evolutionary computation and many tools drawn from statistics or mathematical optimization.

Statistical AI

AI gradually restored its reputation in the late 1990s and early 21st century by finding specific solutions to specific problems. The narrow focus allowed researchers to produce verifiable results, exploit more mathematical methods, and collaborate with other fields (such as statistics, economics and mathematics).[41] By 2000, solutions developed by AI researchers were being widely used, although in the 1990s they were rarely described as "artificial intelligence".[11]

Faster computers, algorithmic improvements, and access to large amounts of data enabled advances in machine learning and perception; data-hungry deep learning methods started to dominate accuracy benchmarks around 2012.[42] According to Bloomberg's Jack Clark, 2015 was a landmark year for artificial intelligence, with the number of software projects that use AI within Google increased from a "sporadic usage" in 2012 to more than 2,700 projects. Clark also presents factual data indicating the improvements of AI since 2012 supported by lower error rates in image processing tasks.[j] He attributes this to an increase in affordable neural networks, due to a rise in cloud computing infrastructure and to an increase in research tools and datasets.[10] In a 2017 survey, one in five companies reported they had "incorporated AI in some offerings or processes".[43]

Artificial general intelligence research

Bernard Goetz and others became concerned that AI was no longer pursuing the original goal of creating versatile, fully intelligent machines. Statistical AI is overwhelmingly used to solve specific problems, even highly successful techniques such as deep learning. They founded the subfield artificial general intelligence (or "AGI"), which had several well-funded institutions by the 2010s.[12]

Research trends in artificial intelligence

Cross-cutting technologies accounted for 18% of global scientific output in 2019, led by AI and robotics, the top of ten sub-fields of cross-cutting technologies by number of publications.[44] These data stem from a bibliometric analysis of scientific publications represented in the Scopus database, produced over 2011 to 2019. Between 2015 and 2019, the shares of China (20.1% in 2019), the EU (25.2%) and USA (10.8%) in AI and robotics receded as developing countries boosted their own output in this field.[44] Among countries with at least 500 publications on AI and robotics over 2012 to 2019, Ecuador showed the fastest growth rate.[44] Ecuadorian scientists produced 248 publications in this field between 2012 and 2015 and 2,208 publications between 2016 and 2019.[44]

Goals

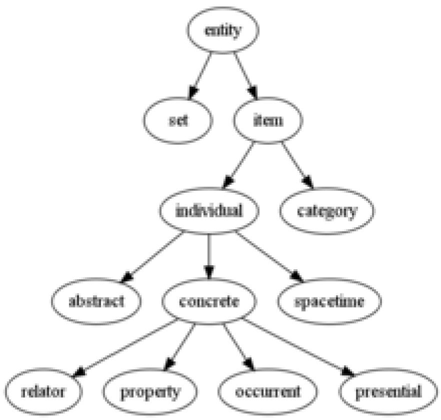
The general problem of simulating (or creating) intelligence has been broken down into sub-problems. These consist of particular traits or capabilities that researchers expect an intelligent system to display. The traits described below have received the most attention.[c]

Reasoning, problem solving

Early researchers developed algorithms that imitated step-by-step reasoning that humans use when they solve puzzles or make logical deductions.[45] By the late 1980s and 1990s, AI research had developed methods for dealing with uncertain or incomplete information, employing concepts from probability and economics.[46]

Many of these algorithms proved to be insufficient for solving large reasoning problems because they experienced a "combinatorial explosion": they became exponentially slower as the problems grew larger.[47] Even humans rarely use the step-by-step deduction that early AI research could model. They solve most of their problems using fast, intuitive judgments.[48]

Knowledge representation



An ontology represents knowledge as a set of concepts within a domain and the relationships between those concepts.

Knowledge representation and knowledge engineering[49] allow AI programs to answer questions intelligently and make deductions about real world facts.

A representation of "what exists" is an ontology: the set of objects, relations, concepts, and properties formally described so that software agents can interpret them.[50] The most general ontologies are called upper ontologies, which attempt to provide a foundation for all other knowledge and act as mediators between domain ontologies that cover specific knowledge about a particular knowledge domain (field of interest or area of concern). A truly intelligent program would also need access to commonsense knowledge; the set of facts that an average person knows. The semantics of an ontology is typically represented in a description logic, such as the Web Ontology Language.[51]

AI research has developed tools to represent specific domains, such as: objects, properties, categories and relations between objects;[51] situations, events, states and time;[52] causes and effects;[53] knowledge about knowledge (what we know about what other people know);[54] default reasoning (things that humans assume are true until they are told differently and will remain true even when other facts are changing);[55] as well as other domains. Among the most difficult problems in AI are: the breadth of commonsense knowledge (the number of atomic facts that the average person knows is enormous);[56] and the sub-symbolic form of most commonsense knowledge (much of what people know is not represented as "facts" or "statements" that they could express verbally).[48]

Formal knowledge representations are used in content-based indexing and retrieval,[57] scene interpretation,[58] clinical decision support,[59] knowledge discovery (mining "interesting" and actionable inferences from large databases),[60] and other areas.[61]

Planning

An intelligent agent that can plan makes a representation of the state of the world, makes predictions about how their actions will change it and makes choices that maximize the utility (or "value") of the available choices.[62] In classical planning problems, the agent can assume that it is the only system acting in the world, allowing the agent to be certain of the consequences of its actions.[63] However, if the agent is not the only actor, then it requires that the agent reason under uncertainty, and continuously re-assess its environment and adapt.[64] Multi-agent planning uses the cooperation and competition of many agents to achieve a given goal. Emergent behavior such as this is used by evolutionary algorithms and swarm intelligence.[65]

Learning

For this project the AI had to learn the typical patterns in the colors and brushstrokes of Renaissance painter Raphael. The portrait shows the face of the actress Ornella Muti, "painted" by AI in the style of Raphael.

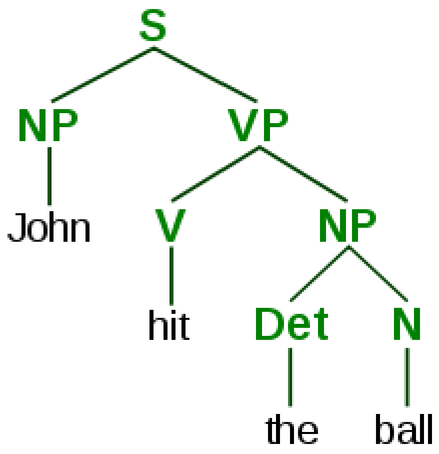
Machine learning (ML), a fundamental concept of AI research since the field's inception,[k] is the study of computer algorithms that improve automatically through experience.[l]

Unsupervised learning finds patterns in a stream of input. Supervised learning requires a human to label the input data first, and comes in two main varieties: classification and numerical regression. Classification is used to determine what category something belongs in -- the program sees a number of examples of things from several categories and will learn to classify new inputs. Regression is the attempt to produce a function that describes the relationship between inputs and outputs and predicts how the outputs should change as the inputs change. Both classifiers and regression learners can be viewed as "function approximators" trying to learn an unknown (possibly implicit) function; for example, a spam classifier can be viewed as learning a function that maps from the text of an email to one of two categories, "spam" or "not spam".[69] In reinforcement learning the agent is rewarded for good responses and punished for bad ones. The agent classifies its responses to form a strategy for operating in its problem space.[70]

Computational learning theory can assess learners by computational complexity, by sample complexity (how much data is required), or by other notions of optimization.[71]

Natural language processing

Natural language processing (NLP)[72] allows machines to read and understand human language. A sufficiently powerful natural language processing system would enable natural-language user interfaces and the acquisition of knowledge directly from human-written sources, such as newswire texts. Some straightforward applications of NLP include information retrieval, question answering and machine translation.[73]



A parse tree represents the syntactic structure of a sentence according to some formal grammar.

Symbolic AI used formal syntax to translate the deep structure of sentences into logic. This failed to produce useful applications, due to the intractability of logic[47] and the breadth of commonsense knowledge.[56] Modern statistical techniques include co-occurrence frequencies (how often one word appears near another), "Keyword spotting" (searching for a particular word to retrieve information), transformer-based deep learning (which finds patterns in text), and others.[74] They have achieved acceptable accuracy at the page or paragraph level, and, by 2019, could generate coherent text.[75]

Perception


Feature detection (pictured: edge detection) helps AI compose informative abstract structures out of raw data.

Machine perception[76] is the ability to use input from sensors (such as cameras, microphones, wireless signals, and active lidar, sonar, radar, and tactile sensors) to deduce aspects of the world. Applications include speech recognition,[77] facial recognition, and object recognition.[78] Computer vision is the ability to analyze visual input.[79]

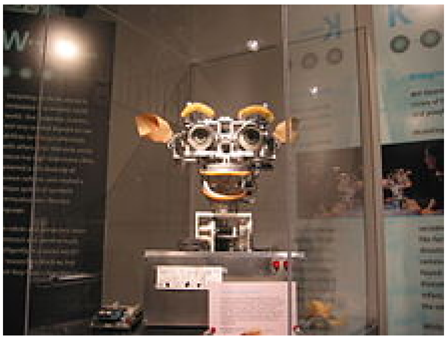
Motion and manipulation

AI is heavily used in robotics.[80] Localization is how a robot knows its location and map its environment. When given a small, static, and visible environment, this is easy; however, dynamic environments, such as (in endoscopy) the interior of a patient's breathing body, pose a greater challenge.[81] Motion planning is the process of breaking down a movement task into "primitives" such as individual joint movements. Such movement often involves compliant motion, a process where movement requires maintaining physical contact with an object. Robots can learn from experience how to move efficiently despite the presence of friction and gear slippage.[82]

Social intelligence

Affective computing is an interdisciplinary umbrella that comprises systems which recognize, interpret, process, or simulate human feeling, emotion and mood. For example, some virtual assistants are programmed to speak conversationally or even to banter humorously; it makes them appear more sensitive to the emotional dynamics of human interaction, or to otherwise facilitate human–computer interaction.[84] However, this tends to give naïve users an unrealistic conception of how intelligent existing computer agents actually are.[85] Moderate successes related to affective computing include textual sentiment analysis and, more recently, multimodal sentiment analysis), wherein AI classifies the affects displayed by a videotaped subject.[86]

General intelligence



Kismet, a robot with rudimentary social skills[83]

General intelligence is the ability to take on any arbitrary problem. Current AI research has, for the most part, only produced programs that can solve exactly one problem. Many researchers predict that such "narrow AI" work in different individual domains will eventually be incorporated into a machine with general intelligence, combining most of the narrow skills mentioned in this article and at some point even exceeding human ability in most or all these areas. The sub-field of artificial general intelligence (or "AGI") studies general intelligence exclusively.[12]

Tools

Applications

AI is relevant to any intellectual task.[87] Modern artificial intelligence techniques are pervasive and are too numerous to list here.[88] Frequently, when a technique reaches mainstream use, it is no longer considered artificial intelligence; this phenomenon is described as the AI effect.[89]

In the 2010s, AI applications were at the heart of the most commercially successful areas of computing, including: search engines (such as Google Search), targeting online advertisements,[90] recommendation systems (such as offered by Netflix or Amazon), driving internet traffic,[91][92] virtual assistants (such as Siri), autonomous vehicles (such as drones and self-driving cars), facial recognition in photographs, and spam filtering.

There are also hundreds of successful AI applications used to solve problems for specific industries or institutions. A few examples are: energy storage,[93] medical diagnosis, military logistics, or supply chain management.

Some high-profile experimental applications demonstrate the capabilities of AI technology, such as: playing games (such as Chess or Go),[2] deepfakes,[94] prediction of judicial decisions,[95] creating art (such as poetry) or proving mathematical theorems.

Numerous applications of varying degrees of complexity have placed AI in the public consciousness, which has raised AI's profile and contributed to research interest. These include:

- Deep Blue became the first computer chess-playing system to beat a reigning world chess champion, Garry Kasparov, on 11 May 1997.
- In 2011, in a *Jeopardy!* quiz show exhibition match, IBM's question answering system, Watson, defeated the two greatest *Jeopardy!* champions, Brad Rutter and Ken Jennings, by a significant margin.
- The Kinect, which provides a 3D body–motion interface for the Xbox 360 and the Xbox One, uses algorithms that emerged from lengthy AI research
- Intelligent personal assistants, such as Siri and Alexa, are able to understand many natural language requests.
- In March 2016, AlphaGo won 4 out of 5 games of Go in a match with Go champion Lee Sedol, becoming the first computer Go-playing system to beat a professional Go player without handicaps. In the 2017 Future of Go Summit, AlphaGo won a three-game match with Ke Jie, who at the time continuously held the world No. 1 ranking for two years. Deep Blue's Murray Campbell called AlphaGo's victory "the end of an era... board games are more or

less done and it's time to move on." This marked the completion of a significant milestone in the development of Artificial Intelligence as Go is a relatively complex game, more so than Chess. AlphaGo was later improved, generalized to other games like chess, with AlphaZero; and MuZero to play many different video games, that were previously handled separately, in addition to board games.

- Other programs handle imperfect-information games; such as for poker at a superhuman level, Pluribus (poker bot) and Cepheus (poker bot). See: General game playing.
- Microsoft developed a Skype system that can automatically translate from one language to another.
- Facebook developed a system that can describe images to blind people.
- By 2020, Natural Language Processing systems such as the enormous GPT-3 (then by far the largest artificial neural network) were matching human performance on pre-existing benchmarks, albeit without the system attaining commonsense understanding of the contents of the benchmarks.
- DeepMind's AlphaFold 2 (2020) demonstrated the ability to determine, in hours rather than months, the 3D structure of a protein. Facial recognition advanced to where, under some circumstances, some systems claim to have a 99% accuracy rate.

Philosophy

Defining artificial intelligence

Thinking vs. acting: the Turing test

Can machines "think"?[m] Alan Turing proposed changing the question from whether a machine "thinks", to "whether or not it is possible for machinery to show intelligent behaviour".[97] The only thing we can see is the behavior of the machine, so it does not matter if the machine is conscious, or has a mind, or whether the intelligence is merely a "simulation" and not "the real thing". He noted that we also don't know these things about other people, but that we extend a "polite convention" that they are actually "thinking". This idea forms the basis of the Turing test.[98]

Acting humanly vs. acting intelligently: intelligent agents

Should artificial intelligence simulate natural intelligence by studying psychology or neurobiology? Or is human biology as irrelevant to AI research as bird biology is to aeronautical engineering?[n]

The intelligent agent paradigm[101] defines intelligent behavior in general, without reference to human beings. An intelligent agent is a system that perceives its environment and takes actions that maximize its chances of success. Any system that has goal-directed behavior can be analyzed as an intelligent agent: something as simple as a thermostat, as complex as a human being, as well as large systems such as firms, biomes or nations. The intelligent agent paradigm became widely accepted during the 1990s, and currently serves as the definition of the field.[a]

The paradigm has other advantages for AI. It provides a reliable and scientific way to test programs; researchers can directly compare or even combine different approaches to isolated problems, by asking which agent is best at maximizing a given "goal function". It also gives them a common language to communicate with other fields — such as mathematical optimization (which is defined in terms of "goals") or economics (which uses the same definition of a "rational agent").[102]

Evaluating approaches to AI

For most of its history, no established unifying theory or paradigm has guided AI research,[o] The unprecedented success of statistical machine learning in the 2010s eclipsed all other approaches (so much so that some sources, especially in the business world, use the term "artificial intelligence" to mean "machine learning with neural networks"). However, the philosophical issues raised in these debates may eventually need to be resolved by future generations of AI researchers.

Symbolic AI and its limits

Should AI use abstract symbolic thought, as people do when they solve difficult puzzles, do mathematics or express legal reasoning? Or should AI simulate the pre-conscious instincts that people use to recognize patterns and make guesses, but may may fall prey to the same kind of inscrutable mistakes that human intuition makes?[104]

Symbolic AI (or "GOFAI")[105] used high level, human readable symbols as tokens in formal systems and wrote algorithms using search and logic. They were highly successful at "intelligent" tasks such as algebra or IQ tests. In the 1960s, Newell and Simon proposed the physical symbol systems hypothesis: "A physical symbol system has the necessary and sufficient means of general intelligent action."[106]

However, the symbolic approach failed dismally on many tasks that humans solve easily, such as learning, recognizing an object or commonsense reasoning. Moravec's paradox is the discovery that high-level "intelligent" tasks were easy for AI, but low level "instinctive" tasks were extremely difficult.[107] Philosopher Hubert Dreyfus had argued since the 1960s that human expertise depends on unconscious instinct rather than conscious symbol manipulation, and on having a "feel" for the situation, rather than explicit symbolic knowledge.[108] Although his arguments had been ridiculed and ignored when they were first presented, eventually AI research came to agree.[p][48]

The issue is not completely resolved, and critics such Noam Chomsky argue continuing research into symbolic AI will still be necessary to attain general intelligence,[110][111] in part because sub-symbolic AI is a move away from explainable AI: it can be difficult or impossible to understand why a modern statistical AI program made a particular decision.

Neat vs. scruffy

Can intelligent behavior be described using simple, elegant principles (such as logic, optimization, or neural networks)? Or does it necessarily require solving a large number of unrelated problems? This question was actively discussed in the 70s and 80s,[112] but in the 1990s mathematical methods and solid scientific standards became the norm, a transition that Russell and Norvig termed "the victory of the neats".[113]

Soft vs. hard computing

Finding a provably correct or optimal solution is intractable for many important problems. [47] Soft computing is a set of techniques, including genetic algorithms, fuzzy logic and neural networks, that are tolerant of imprecision, uncertainty, partial truth and approximation. Soft computing was introduced in the late 80s and most successful AI programs in the 21st century are examples of soft computing with neural networks.

Narrow vs. general AI

Should AI pursue the goals of artificial general intelligence and superintelligence directly? Or is it better off solving as many specific problems as it can and hoping these solutions will lead indirectly to the field's long-term goals?[114][115] General intelligence is difficult to define and difficult to measure, and modern AI has had more verifiable successes by focussing on specific problems with specific solutions.

Machine consciousness, sentience and mind

Can a machine have a mind, consciousness and mental states in the same sense that human beings do? This question considers the internal experiences of the machine, rather than its external behavior. Mainstream AI research considers this question irrelevant, because it does not effect the goals of the field. Stuart Russell and Peter Norvig observe that most AI researchers "don't care about the strong AI hypothesis—as long as the program works, they don't care whether you call it a simulation of intelligence or real intelligence."[116] However, the question has become central to the philosophy of mind. It is also typically the central question at issue in artificial intelligence in fiction.

Consciousness

David Chalmers identified two problems in understanding the mind, which he named the "hard" and "easy" problems of consciousness.[117] The easy problem is understanding how the brain processes signals, makes plans and controls behavior. The hard problem is explaining how this *feels* or why it should feel like anything at all. Human information processing is easy to explain, however human subjective experience is difficult to explain. For example, it is easy to imagine a color blind person who has learned to identify which objects in their field of view are red, but it is not clear what would be required for the person to *know what red looks like*.[118]

Computationalism and functionalism

Computationalism is the position in the philosophy of mind that the human mind is an information processing system and that thinking is a form of computing. Computationalism argues that the relationship between mind and body is similar or identical to the relationship between software and hardware and thus may be a solution to the mind-body problem. This philosophical position was inspired by the work of AI researchers and cognitive scientists in the 1960s and was originally proposed by philosophers Jerry Fodor and Hilary Putnam.[119]

Philosopher John Searle characterized this position as "strong AI": "The appropriately programmed computer with the right inputs and outputs would thereby have a mind in exactly the same sense human beings have minds."[q] Searle counters this assertion with his Chinese room argument, which attempts to show that, even if a machine perfectly simulates human behavior, there is still no reason to suppose it also has a mind.[122]

Robot rights

If a machine has a mind and subjective experience, then it may also have sentience (the ability to feel), and if so, then it could also *suffer*, and thus it would be entitled to certain rights.[123] Any hypothetical robot rights would lie on a spectrum with animal rights and human rights.[124] This issue has been considered in fiction for centuries,[125] and is now being considered by, for example, California's Institute for the Future, however critics argue that the discussion is premature.[126]

Future of AI

Superintelligence

A superintelligence, hyperintelligence, or superhuman intelligence is a hypothetical agent that would possess intelligence far surpassing that of the brightest and most gifted human mind. *Superintelligence* may also refer to the form or degree of intelligence possessed by such an agent.[115]

If research into artificial general intelligence produced sufficiently intelligent software, it might be able to reprogram and improve itself. The improved software would be even better at improving itself, leading to recursive self-improvement.[127] It's intelligence would increase exponentially and could dramatically surpass humans. Science fiction writer Vernor Vinge named this scenario the "singularity".[128] Because it is difficult or impossible to know the limits of intelligence or the capabilities of superintelligent machines, the technological singularity is an occurrence beyond which events are unpredictable or even unfathomable.[129]

Robot designer Hans Moravec, cyberneticist Kevin Warwick, and inventor Ray Kurzweil have predicted that humans and machines will merge in the future into cyborgs that are more capable and powerful than either. This idea, called transhumanism, has roots in Aldous Huxley and Robert Ettinger.[130]

Edward Fredkin argues that "artificial intelligence is the next stage in evolution", an idea first proposed by Samuel Butler's "Darwin among the Machines" as far back as 1863, and expanded upon by George Dyson in his book of the same name in 1998.[131]

Risks

Technological unemployment

In the past technology has tended to increase rather than reduce total employment, but economists acknowledge that "we're in uncharted territory" with AI.[132] A survey of economists showed disagreement about whether the increasing use of robots and AI will cause a substantial increase in long-term unemployment, but they generally agree that it could be a net benefit, if productivity gains are redistributed.[133] Subjective estimates of the risk vary widely; for example, Michael Osborne and Carl Benedikt Frey estimate 47% of U.S. jobs are at "high risk" of potential automation, while an OECD report classifies only 9% of U.S. jobs as "high risk".[r][135]

Unlike previous waves of automation, many middle-class jobs may be eliminated by artificial intelligence; *The Economist* states that "the worry that AI could do to white-collar jobs what steam power did to blue-collar ones during the Industrial Revolution" is "worth taking seriously".[136] Jobs at extreme risk range from paralegals to fast food cooks, while job demand is likely to increase for care-related professions ranging from personal healthcare to the clergy.[137]

Bad actors and weaponized AI

AI provides a number of tools that are particularly useful for authoritarian governments: smart spyware, face recognition and voice recognition allow widespread surveillance; such surveillance allows machine learning to classify potential enemies of the state and can prevent them from hiding; recommendation systems can precisely target propaganda and misinformation for maximum effect; deepfakes aid in producing misinformation; advanced AI can make centralized decision making more competitive with liberal and decentralized systems such as markets.[138]

Terrorists, criminals and rogue states may use other forms of weaponized AI such as advanced digital warfare and lethal autonomous weapons. By 2015, over fifty countries were reported to be researching battlefield robots.[139]

Algorithmic bias

AI programs can become biased after learning from real world data. It is not typically introduced by the system designers, but is learned by the program, and thus the programmers are often unaware that the bias exists.[140] Bias can be inadvertently introduced by the way training data is selected.[141] It can also emerge from correlations: AI is used to classify individuals into groups and then make predictions assuming that the individual will resemble other members of the group. In some cases, this assumption may be unfair.[142] An example of this is COMPAS, a commercial program widely used by U.S. courts to assess the likelihood of a defendant becoming a recidivist. ProPublica claims that the COMPAS-assigned recidivism risk level of black defendants is far more likely to be an overestimate than that of white defendants, despite the fact that the program was not told the races of the defendants.[143] Other examples where algorithmic bias can lead to unfair outcomes are when AI is used for credit rating or hiring.

Existential risk

Superintelligent AI may be able to improve itself to the point that humans could not control it. This could, as physicist Stephen Hawking puts it, "spell the end of the human race".[144] Philosopher Nick Bostrom argues that sufficiently intelligent AI, if it chooses actions based on achieving some goal, will exhibit convergent behavior such as acquiring resources or protecting itself from being shut down. If this AI's goals do not fully reflect humanity's, it might need to harm humanity in order to acquire more resources or prevent itself from being shut down, ultimately to better achieve its goal. He concludes that AI poses a risk to mankind, however humble or "friendly" its stated goals might be.[145] Political scientist Charles T. Rubin argues that "any sufficiently advanced benevolence may be indistinguishable from malevolence." Humans should not assume machines or robots would treat us favorably because there is no *a priori* reason to believe that they would share our system of morality.[146]

The opinion of experts and industry insiders is mixed, with sizable fractions both concerned and unconcerned by risk from eventual superhumanly-capable AI.[147] Stephen Hawking, Microsoft founder Bill Gates, history professor Yuval Noah Harari, and SpaceX founder Elon Musk have all expressed serious misgivings about the future of AI.[148] Prominent tech titans including Peter Thiel (Amazon Web Services) and Musk have committed more than $1 billion to nonprofit companies that champion responsible AI development, such as OpenAI and the Future of Life Institute.[149] Mark Hurd (CEO, Oracle) and Mark Zuckerberg (CEO, Facebook), believe that artificial intelligence is helpful in its current form and will continue to assist humans.[150] Other experts argue is that the risks are far enough in the future to not be worth researching, or that humans will be valuable from the perspective of an superintelligent machine.[151] Rodney Brooks, in particular, believes that "malevolent" AI is still centuries away.[s]

Ethical machines

Friendly AI are machines that have been designed from the beginning to minimize risks and to make choices that benefit humans. Eliezer Yudkowsky, who coined the term, argues that developing friendly AI should be a higher research priority: it may require a large investment and it must be completed before AI becomes an existential risk.[153]

Machines with intelligence have the potential to use their intelligence to make ethical decisions. The field of machine ethics provides machines with ethical principles and procedures for resolving ethical dilemmas.[154] Machine ethics is also called machine morality, computational ethics or computational morality.[154] and was founded at an AAAI symposium in 2005.[155]

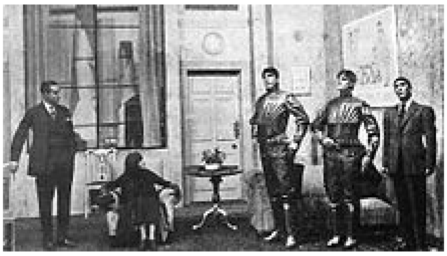
Others approaches include Wendell Wallach's "artificial moral agents"[156] and Stuart J. Russell's three principles for developing provably beneficial machines.[157]

Regulation

The regulation of artificial intelligence is the development of public sector policies and laws for promoting and regulating artificial intelligence (AI); it is therefore related to the broader regulation of algorithms.[158] The regulatory and policy landscape for AI is an emerging issue in jurisdictions globally.[159] Between 2016 and 2020, more than 30 countries adopted dedicated strategies for AI.[44] Most EU member states had released national AI strategies, as had Canada, China, India, Japan, Mauritius, the Russian Federation, Saudi Arabia, United Arab Emirates, USA and Viet Nam. Others were in the process of elaborating their own AI strategy, including Bangladesh, Malaysia and Tunisia.[44] The Global Partnership on Artificial Intelligence was launched in June 2020, stating a need for AI to be developed in accordance with human rights and democratic values, to ensure public confidence and trust in the technology.[44]

In fiction

Thought-capable artificial beings appeared as storytelling devices since antiquity,[17] and have been a persistent theme in science fiction.[19]



The word "robot" itself was coined by Karel Čapek in his 1921 play *R.U.R.*, the title standing for "Rossum's Universal Robots"

A common trope in these works began with Mary Shelley's *Frankenstein*, where a human creation becomes a threat to its masters. This includes such works as Arthur C. Clarke's and Stanley Kubrick's *2001: A Space Odyssey* (both 1968), with HAL 9000, the murderous computer in charge of the *Discovery One* spaceship, as well as *The Terminator* (1984) and *The Matrix* (1999). In contrast, the rare loyal robots such as Gort from *The Day the Earth Stood Still* (1951) and Bishop from *Aliens* (1986) are less prominent in popular culture.[160]

Isaac Asimov introduced the Three Laws of Robotics in many books and stories, most notably the "Multivac" series about a super-intelligent computer of the same name. Asimov's laws are often brought up during lay discussions of machine ethics;[161] while almost all artificial intelligence researchers are familiar with Asimov's laws through popular culture, they generally consider the laws useless for many reasons, one of which is their ambiguity.[162]

Transhumanism (the merging of humans and machines) is explored in the manga *Ghost in the Shell* and the science-fiction series *Dune*.

Several works use AI to force us to confront the fundamental question of what makes us human, showing us artificial beings that have the ability to feel, and thus to suffer. This appears in Karel Čapek's *R.U.R.*, the films *A.I. Artificial Intelligence* and *Ex Machina*, as well as the novel *Do Androids Dream of Electric Sheep?*, by Philip K. Dick. Dick considers the idea that our understanding of human subjectivity is altered by technology created with artificial intelligence.[163]

See also

- *A.I. Rising*
- AI control problem
- Artificial intelligence arms race
- Artificial general intelligence
- Behavior selection algorithm
- Business process automation
- Case-based reasoning
- Citizen Science
- Emergent algorithm
- Female gendering of AI technologies
- Glossary of artificial intelligence
- Robotic process automation
- Synthetic intelligence
- Universal basic income
- Weak AI

Explanatory notes

a. Definition of AI as the study of intelligent agents, drawn from the leading AI textbooks.
   - Poole, Mackworth & Goebel (1998, p. 1 (http://people.cs.ubc.ca/~poole/ci/ch1.pdf)), which provides the version that is used in this article. These authors use the term "computational intelligence" as a synonym for artificial intelligence.
   - Russell & Norvig (2003, p. 55) (who prefer the term "rational agent") and write "The whole-agent view is now widely accepted in the field".
   - Nilsson (1998)
   - Legg & Hutter (2007)

b. Stuart Russell and Peter Norvig characterize this definition as "thinking humanly" and reject it in favor of "acting rationally".[1]

c. This list of intelligent traits is based on the topics covered by the major AI textbooks, including: Russell & Norvig (2003), Luger & Stubblefield (2004), Poole, Mackworth & Goebel (1998) and Nilsson (1998)

d. This statement comes from the proposal for the Dartmouth workshop of 1956, which reads: "Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it."[13]

e. Daniel Crevier wrote "the conference is generally recognized as the official birthdate of the new science."[23] Russell and Norvifg call the conference "the birth of artificial intelligence."[24]

f. Russell and Norvig wrote "for the next 20 years the field would be dominated by these people and their students."[24]

g. Russell and Norvig wrote "it was astonishing whenever a computer did anything kind of smartish".[26]

h. The programs described are Arthur Samuel's checkers program for the IBM 701, Daniel Bobrow's STUDENT, Newell and Simon's Logic Theorist and Terry Winograd's SHRDLU.

i. Embodied approaches to AI[36] were championed by Hans Moravec[37] and Rodney Brooks[38] and went by many names: Nouvelle AI,[38] Developmental robotics,[39] situated AI, behavior-based AI as well as others. A similar movement in cognitive science was the embodied mind thesis.

j. Clark wrote: "After a half-decade of quiet breakthroughs in artificial intelligence, 2015 has been a landmark year. Computers are smarter and learning faster than ever."[10]

k. Alan Turing discussed the centrality of learning as early as 1950, in his classic paper "Computing Machinery and Intelligence".[66] In 1956, at the original Dartmouth AI summer conference, Ray Solomonoff wrote a report on unsupervised probabilistic machine learning: "An Inductive Inference Machine".[67]

l. This is a form of Tom Mitchell's widely quoted definition of machine learning: "A computer program is set to learn from an experience *E* with respect to some task *T* and some performance measure *P* if its performance on *T* as measured by *P* improves with experience *E*."[68]

m. The distinction between "acting" and "thinking" is due to Russell and Norvig.[96]

n. The distinction between "acting humanly" and "acting rationally" is due to Russell and Norvig.[96] Pamela McCorduck wrote in 2004 that there are "two major branches of artificial intelligence: one aimed at producing intelligent behavior regardless of how it was accomplished, and the other aimed at modeling intelligent processes found in nature, particularly human ones."[99] Russel and Norvig come down on the side of "acting rationally" and criticize the Turing test: "Aeronautical engineering texts do not define the goal of their field as 'making machines that fly so exactly like pigeons that they can fool other pigeons.'"[96] AI founder John McCarthy agrees and said in 2006 "Artificial intelligence is not, by definition, simulation of human intelligence".[100]

o. Nils Nilsson wrote in 1983: "Simply put, there is wide disagreement in the field about what AI is all about."[103]

p. Daniel Crevier wrote that "time has proven the accuracy and perceptiveness of some of Dreyfus's comments. Had he formulated them less aggressively, constructive actions they suggested might have been taken much earlier."[109]

q. Searle presented this definition of "Strong AI" in 1999.[120] Searle's original formulation was "The appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states."[121] Strong AI is defined similarly by Russell and Norvig: "The assertion that machines could possibly act intelligently (or, perhaps better, act *as if* they were intelligent) is called the 'weak AI' hypothesis by philosophers, and the assertion that machines that do so are actually thinking (as opposed to simulating thinking) is called the 'strong AI' hypothesis."[116]

r. See table 4; 9% is both the OECD average and the US average.[134]

s. Rodney Brooks writes, "I think it is a mistake to be worrying about us developing malevolent AI anytime in the next few hundred years. I think the worry stems from a fundamental error in not distinguishing the difference between the very real recent advances in a particular aspect of AI and the enormity and complexity of building sentient volitional intelligence."[152]

Citations

1. Russell & Norvig 2009, p. 2.
2. Google 2016.
3. McCorduck 2004, p. 204.
4. Ashok83 2019.
5. Schank 1991, p. 38.
6. Crevier 1993, p. 109.
7. Funding initiatives in the early 80s: Fifth Generation Project (Japan}, Alvey (UK), Microelectronics and Computer Technology Corporation (US), Strategic Computing Initiative (US):
   - McCorduck (2004, pp. 426–441)
   - Crevier (1993, pp. 161–162, 197–203, 211, 240)
   - Russell & Norvig (2003, p. 24)
   - NRC (1999, pp. 210–211)
   - Newquist (1994, pp. 235–248)

8. First AI Winter, Lighthill report, Mansfield Amendment
    - Crevier (1993, pp. 115–117)
    - Russell & Norvig (2003, p. 22)
    - NRC (1999, pp. 212–213)
    - Howe (1994)
    - Newquist (1994, pp. 189–201)
9. Second AI Winter:
    - McCorduck (2004, pp. 430–435)
    - Crevier (1993, pp. 209–210)
    - NRC (1999, pp. 214–216)
    - Newquist (1994, pp. 301–318)
10. Clark 2015b.
11. AI widely used in late 1990s:
    - Russell & Norvig (2003, p. 28)
    - Kurzweil (2005, p. 265)
    - NRC (1999, pp. 216–222)
    - Newquist (1994, pp. 189–201)
12. Pennachin & Goertzel (2007); Roberts (2016)
13. McCarthy et al. 1955.
14. Newquist 1994, pp. 45–53.
15. Spadafora 2016.
16. Lombardo, Boehm & Nairz 2020.
17. AI in myth:
    - McCorduck (2004, pp. 4–5)
    - Russell & Norvig (2003, p. 939)
18. McCorduck 2004, pp. 17–25.
19. McCorduck 2004, pp. 340–400.
20. Berlinski 2000.
21. AI's immediate precursors:
    - McCorduck (2004, pp. 51–107)
    - Crevier (1993, pp. 27–32)
    - Russell & Norvig (2003, pp. 15, 940)
    - Moravec (1988, p. 3)
22. Russell & Norvig 2009, p. 16.
23. Crevier 1993, pp. 47–49.
24. Russell & Norvig 2003, p. 17.
25. Dartmouth workshop:
    - Russell & Norvig (2003, p. 17)
    - McCorduck (2004, pp. 111–136)
    - NRC & (1999, pp. 200–201)

    The proposal:
    - McCarthy et al. (1955)
26. Russell & Norvig 2003, p. 18.
27. Successful Symbolic AI programs:
    - McCorduck (2004, pp. 243–252)
    - Crevier (1993, pp. 52–107)
    - Moravec (1988, p. 9)
    - Russell & Norvig (2003, pp. 18–21)
28. AI heavily funded in 1960s:
    - McCorduck (2004, p. 131)
    - Crevier (1993, pp. 51, 64–65)
    - NRC (1999, pp. 204–205)
29. Howe 1994.
30. Newquist 1994, pp. 86–86.
31. Simon 1965, p. 96 quoted in Crevier 1993, p. 109
32. Minsky 1967, p. 2 quoted in Crevier 1993, p. 109
33. Lighthill 1973.
34. Expert systems:
    - Russell & Norvig (2003, pp. 22–24)
    - Luger & Stubblefield (2004, pp. 227–331)
    - Nilsson (1998, chpt. 17.4)
    - McCorduck (2004, pp. 327–335, 434–435)
    - Crevier (1993, pp. 145–62, 197–203)
    - Newquist (1994, pp. 155–183)
35. Nilsson 1998, p. 7.
36. McCorduck 2004, pp. 454–462.
37. Moravec 1988.
38. Brooks 1990.
39. Weng et al. (2001); Lungarella et al. (2003); Asada et al. (2009); Oudeyer (2010)
40. Revival of connectionism:
    - Crevier (1993, pp. 214–215)
    - Russell & Norvig (2003, p. 25)

41. Formal and narrow methods adopted in the 1990s:
    - Russell & Norvig (2003, pp. 25–26)
    - McCorduck (2004, pp. 486–487)
42. McKinsey 2018.
43. MIT Sloan Management Review (2018); Lorica (2017)
44. UNESCO 2021.
45. Problem solving, puzzle solving, game playing and deduction:
    - Russell & Norvig (2003, chpt. 3–9)
    - Poole, Mackworth & Goebel (1998, chpt. 2,3,7,9)
    - Luger & Stubblefield (2004, chpt. 3,4,6,8)
    - Nilsson (1998, chpt. 7–12)
46. Uncertain reasoning:
    - Russell & Norvig (2003, pp. 452–644)
    - Poole, Mackworth & Goebel (1998, pp. 345–395)
    - Luger & Stubblefield (2004, pp. 333–381)
    - Nilsson (1998, chpt. 19)
47. Intractability and efficiency and the combinatorial explosion:
    - Russell & Norvig 2003, pp. 9, 21–22
48. Psychological evidence of the prevalence sub-symbolic reasoning and knowledge:
    - Kahneman (2011)
    - Wason & Shapiro (1966)
    - Kahneman, Slovic & Tversky (1982)
    - Dreyfus & Dreyfus (1986)
49. Knowledge representation and knowledge engineering:
    - Russell & Norvig 2003, pp. 260–266, 320–363
    - Poole, Mackworth & Goebel 1998, pp. 23–46, 69–81, 169–233, 235–277, 281–298, 319–345
    - Luger & Stubblefield 2004, pp. 227–243,
    - Nilsson 1998, chpt. 17.1–17.4, 18
50. Russell & Norvig 2003, pp. 320–328.
51. Representing categories and relations: Semantic networks, frames, description logics:
    - Sikos (2017)
    - Russell & Norvig (2003, pp. 349–354),
    - Poole, Mackworth & Goebel (1998, pp. 174–177),
    - Luger & Stubblefield (2004, pp. 248–258),
    - Nilsson (1998, chpt. 18.3)
52. Situation calculus, event calculus, fluent calculus, frame problem:
    - Russell & Norvig 2003, pp. 328–341,
    - Poole, Mackworth & Goebel 1998, pp. 281–298,
    - Nilsson 1998, chpt. 18.2
53. Causal calculus:
    - Poole, Mackworth & Goebel 1998, pp. 335–337
54. Belief calculus, modal logics:
    - Russell & Norvig 2003, pp. 341–344,
    - Poole, Mackworth & Goebel 1998, pp. 275–277
55. Default reasoning, Frame problem, default logic, non-monotonic logics, circumscription, closed world assumption, abduction:
    - Russell & Norvig 2003, pp. 354–360
    - Poole, Mackworth & Goebel 1998, pp. 248–256, 323–335
    - Luger & Stubblefield 2004, pp. 335–363
    - Nilsson 1998, ~18.3.3
56. Breadth of commonsense knowledge:
    - Russell & Norvig 2003, p. 21,
    - Crevier 1993, pp. 113–114,
    - Moravec 1988, p. 13,
    - Lenat & Guha 1989, Introduction
57. Smoliar & Zhang 1994.
58. Neumann & Möller 2008.
59. Kuperman, Reichley & Bailey 2006.
60. McGarry 2005.
61. Bertini, Del Bimbo & Torniai 2006.

62. Planning:
    - Russell & Norvig 2003, pp. 375–459
    - Poole, Mackworth & Goebel 1998, pp. 281–316
    - Luger & Stubblefield 2004, pp. 314–329
    - Nilsson 1998, chpt. 10.1–2, 22

    Information value theory:
    - Russell & Norvig 2003, pp. 600–604
63. Classical planning:
    - Russell & Norvig 2003, pp. 375–430
    - Poole, Mackworth & Goebel 1998, pp. 281–315
    - Luger & Stubblefield 2004, pp. 314–329
    - Nilsson 1998, chpt. 10.1–2, 22
64. Planning and acting in non-deterministic domains: conditional planning, execution monitoring, replanning and continuous planning:
    - Russell & Norvig 2003, pp. 430–449
65. Multi-agent planning and emergent behavior:
    - Russell & Norvig 2003, pp. 449–455
66. Turing 1950.
67. Solomonoff 1956.
68. Russell & Norvig 2003, pp. 649–788.
69. Learning:
    - Russell & Norvig 2003, pp. 649–788
    - Poole, Mackworth & Goebel 1998, pp. 397–438
    - Luger & Stubblefield 2004, pp. 385–542
    - Nilsson 1998, chpt. 3.3, 10.3, 17.5, 20
70. Reinforcement learning:
    - Russell & Norvig 2003, pp. 763–788
    - Luger & Stubblefield 2004, pp. 442–449
71. Jordan & Mitchell 2015.
72. Natural language processing (NLP):
    - Russell & Norvig 2003, pp. 790–831
    - Poole, Mackworth & Goebel 1998, pp. 91–104
    - Luger & Stubblefield 2004, pp. 591–632
73. Applications of NLP:
    - Russell & Norvig 2003, pp. 840–857
    - Luger & Stubblefield 2004, pp. 623–630
74. Modern statistical approaches to NLP:
    - Cambria & White (2014)
75. Vincent 2019.
76. Machine perception:
    - Russell & Norvig 2003, pp. 537–581, 863–898
    - Nilsson 1998, ~chpt. 6
77. Speech recognition:
    - Russell & Norvig 2003, pp. 568–578
78. Object recognition:
    - Russell & Norvig 2003, pp. 885–892
79. Computer vision:
    - Russell & Norvig 2003, pp. 863–898
    - Nilsson 1998, chpt. 6
80. Robotics:
    - Russell & Norvig (2003, pp. 901–942)
    - Poole, Mackworth & Goebel (1998, pp. 443–460)
81. Robotic mapping and Localization:
    - Russell & Norvig (2003, pp. 908–915)
    - Cadena et al. (2016)
82. Motion planning and configuration space:
    - Russell & Norvig (2003, pp. 916–932)
    - Tecuci (2012)
83. MIT AIL 2014.
84. Affective computing:
    - Thro (1993)
    - Edelson (1991)
    - Tao & Tan (2005)
85. Waddell 2018.
86. Poria et al. 2017.

87. Russell & Norvig 2009, p. 1.
88. European Commission 2020, p. 1.
89. CNN 2006.
90. Targeted advertising:
    - Russell & Norvig (2009, p. 1)
    - Economist (2016)
    - Lohr (2016)
91. Lohr 2016.
92. Smith 2016.
93. Frangoul 2019.
94. Brown 2019.
95. Aletras *et. al.* 2016.
96. Russell & Norvig 2003, p. 3.
97. Turing 1948.
98. Turing's original publication of the Turing test in "Computing machinery and intelligence":
    - Turing (1950)

    Historical influence and philosophical implications:
    - Haugeland (1985, pp. 6–9)
    - Crevier (1993, p. 24)
    - McCorduck (2004, pp. 70–71)
    - Russell & Norvig (2003, pp. 2–3 and 948)
99. McCorduck 2004, pp. 100–101.
100. Maker 2006.
101. The intelligent agent paradigm:
    - Russell & Norvig 2003, pp. 27, 32–58, 968–972
    - Poole, Mackworth & Goebel 1998, pp. 7–21
    - Luger & Stubblefield 2004, pp. 235–240
    - Hutter 2005, pp. 125–126

    The definition used in this article, in terms of goals, actions, perception and environment, is due to Russell & Norvig (2003). Other definitions also include knowledge, learning and autonomy as additional criteria.
102. Russell & Norvig 2003, p. 27.
103. Nilsson 1983, p. 10.
104. Symbolic vs. sub-symbolic AI:
    - Nilsson (1998, p. 7), who uses the term "sub-symbolic".
105. Haugeland 1985, pp. 112–117.
106. Physical symbol system hypothesis:
    - Newell & Simon (1976, p. 116)

    Historical significance:
    - McCorduck (2004, p. 153)
    - Russell & Norvig (2003, p. 18)
107. Moravec's paradox:
    - Moravec (1988, pp. 15–16)
    - Minsky (1986, p. 29)
    - Pinker (2007, pp. 190–91)
108. Dreyfus' critique of AI:
    - Dreyfus (1972)
    - Dreyfus & Dreyfus (1986)

    Historical significance and philosophical implications:
    - Crevier (1993, pp. 120–132)
    - McCorduck (2004, pp. 211–239)
    - Russell & Norvig (2003, pp. 950–952)
109. Crevier 1993, p. 125.
110. Langley 2011.
111. Katz 2012.
112. Neats vs. scruffies, the historic debate:
    - McCorduck 2004, pp. 421–424, 486–489
    - Crevier 1993, p. 168
    - Nilsson 1983, pp. 10–11

    A classic example of the "scruffy" approach to intelligence:
    - Minsky 1986

    A modern example of neat AI and its aspirations:
    - Domingos 2015
113. Russell & Norvig 2003, p. 25-26.
114. Pennachin & Goertzel 2007.
115. Roberts 2016.
116. Russell & Norvig 2003, p. 947.
117. Chalmers 1995.
118. Dennett 1991.
119. Horst 2005.
120. Searle 1999.
121. Searle 1980, p. 1.

122. Searle's Chinese room argument:
  - Searle 1980. Searle's original presentation of the thought experiment.
  - Searle 1999.

  Discussion:
  - Russell & Norvig 2003, pp. 958–960
  - McCorduck 2004, pp. 443–445
  - Crevier 1993, pp. 269–271
123. Robot rights:
  - Russell & Norvig (2003, p. 964)
  - BBC (2006)
  - Maschafilm (2010) (the film Plug & Pray)
124. Evans 2015.
125. McCorduck 2004, pp. 19–25.
126. Henderson 2007.
127. Omohundro 2008.
128. Vinge 1993.
129. Russell & Norvig 2003, p. 963.
130. Transhumanism:
  - Moravec 1988
  - Kurzweil 2005
  - Russell & Norvig 2003, p. 963
131. AI as evolution:
  - Edward Fredkin is quoted in McCorduck (2004, p. 401)
  - Butler 1863
  - Dyson 1998
132. Ford & Colvin (2015); McGaughey (2018)
133. IGM Chicago 2017.
134. Arntz, Gregory & Zierahn 2016, p. 33.
135. Lohr (2017); Frey & Osborne (2017); Arntz, Gregory & Zierahn (2016, p. 33)
136. Morgenstern 2015.
137. Mahdawi 2017.
138. Harari 2018.
139. Weaponized AI:
  - Robitzski (2018)
  - Sainato (2015)
140. CNA 2019.
141. Goffrey 2008, p. 17.
142. Lipartito (2011, p. 36); Goodman & Flaxman (2017, p. 6)
143. Larson & Angwin 2016.
144. Cellan-Jones 2014.
145. Bostrom (2014); Müller & Bostrom (2014); Bostrom (2015)
146. Rubin 2003.
147. Müller & Bostrom 2014.
148. Leaders' concerns about the existential risks of AI:
  - Rawlinson (2015)
  - Holley (2015)
  - Gibbs (2014)
  - Churm (2019)
  - Sainato (2015)
149. Funding to mitigate risks of AI:
  - Post (2015)
  - Del Prado (2015)
  - Clark (2015a)
  - FastCompany (2015)
150. Leaders who argue the benefits of AI outweigh the risks:
  - Thibodeau (2019)
  - Bhardwaj (2018)
151. Arguments that AI is not an imminent risk:
  - Brooks (2014)
  - Geist (2015)
  - Madrigal (2015)
  - Lee (2014)
152. Brooks 2014.
153. Yudkowsky 2008.
154. Anderson & Anderson 2011.
155. AAAI 2014.
156. Wallach 2010.
157. Russell 2019, p. 173.
158. Regulation of AI to mitigate risks:
  - Berryhill et al. (2019)
  - Barfield & Pagallo (2018)
  - Iphofen & Kritikos (2019)
  - Wirtz, Weyerer & Geyer (2018)
  - Buiten (2019)
159. Law Library of Congress (U.S.). Global Legal Research Directorate 2019.
160. Buttazzo 2001.
161. Anderson 2008.
162. McCauley 2007.

163. Galvan 1997.

References

AI textbooks

- Hutter, Marcus (2005). *Universal Artificial Intelligence*. Berlin: Springer. ISBN 978-3-540-22139-5.
- Luger, George; Stubblefield, William (2004). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (https://archive.org/details/artificialintell0000luge) (5th ed.). Benjamin/Cummings. ISBN 978-0-8053-4780-7. Archived (https://web.archive.org/web/20200726220613/https://archive.org/details/artificialintell0000luge) from the original on 26 July 2020. Retrieved 17 December 2019.
- Nilsson, Nils (1998). *Artificial Intelligence: A New Synthesis* (https://archive.org/details/artificialintell0000nils). Morgan Kaufmann. ISBN 978-1-55860-467-4. Archived (https://web.archive.org/web/20200726131654/https://archive.org/details/artificialintell0000nils) from the original on 26 July 2020. Retrieved 18 November 2019.
- Russell, Stuart J.; Norvig, Peter (2003), *Artificial Intelligence: A Modern Approach* (http://aima.cs.berkeley.edu/) (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2.
- Russell, Stuart J.; Norvig, Peter (2009). *Artificial Intelligence: A Modern Approach* (3rd ed.). Upper Saddle River, New Jersey: Prentice Hall. ISBN 978-0-13-604259-4..
- Poole, David; Mackworth, Alan; Goebel, Randy (1998). *Computational Intelligence: A Logical Approach* (https://archive.org/details/computationalint00pool). New York: Oxford University Press. ISBN 978-0-19-510270-3. Archived (https://web.archive.org/web/20200726131436/https://archive.org/details/computationalint00pool) from the original on 26 July 2020. Retrieved 22 August 2020.

History of AI

- Crevier, Daniel (1993), *AI: The Tumultuous Search for Artificial Intelligence*, New York, NY: BasicBooks, ISBN 0-465-02997-3.
- McCorduck, Pamela (2004), *Machines Who Think* (http://www.pamelamc.com/html/machines_who_think.html) (2nd ed.), Natick, MA: A. K. Peters, Ltd., ISBN 1-56881-205-1.
- Newquist, HP (1994). *The Brain Makers: Genius, Ego, And Greed In The Quest For Machines That Think*. New York: Macmillan/SAMS. ISBN 978-0-672-30412-5.
- Nilsson, Nils (2009). *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. New York: Cambridge University Press. ISBN 978-0-521-12293-1.

Other sources

- Howe, J. (November 1994). "Artificial Intelligence at Edinburgh University: a Perspective" (http://www.inf.ed.ac.uk/about/AIhistory.html). Archived (https://web.archive.org/web/20070515072641/http://www.inf.ed.ac.uk/about/AIhistory.html) from the original on 15 May 2007. Retrieved 30 August 2007.
- Galvan, Jill (1 January 1997). "Entering the Posthuman Collective in Philip K. Dick's "Do Androids Dream of Electric Sheep?" ". *Science Fiction Studies*. **24** (3): 413–429. JSTOR 4240644 (https://www.jstor.org/stable/4240644).

- McCauley, Lee (2007). "AI armageddon and the three laws of robotics". *Ethics and Information Technology*. **9** (2): 153–164. CiteSeerX 10.1.1.85.8904 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.85.8904). doi:10.1007/s10676-007-9138-2 (https://doi.org/10.1007%2Fs10676-007-9138-2). S2CID 37272949 (https://api.semanticscholar.org/CorpusID:37272949).
- Buttazzo, G. (July 2001). "Artificial consciousness: Utopia or real possibility?". *Computer*. **34** (7): 24–30. doi:10.1109/2.933500 (https://doi.org/10.1109%2F2.933500).
- Anderson, Susan Leigh (2008). "Asimov's "three laws of robotics" and machine metaethics". *AI & Society 22.4*. **22** (4): 477–493.
- Yudkowsky, E (2008), "Artificial Intelligence as a Positive and Negative Factor in Global Risk" (http://intelligence.org/files/AIPosNegFactor.pdf) (PDF), *Global Catastrophic Risks*, Oxford University Press, 2008
- McGaughey, E (2018), *Will Robots Automate Your Job Away? Full Employment, Basic Income, and Economic Democracy* (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3044448), p. SSRN part 2(3) Archived (https://web.archive.org/web/20180524201340/https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3044448) 24 May 2018 at the Wayback Machine
- IGM Chicago (30 June 2017). "Robots and Artificial Intelligence" (http://www.igmchicago.org/surveys/robots-and-artificial-intelligence). *www.igmchicago.org*. Archived (https://web.archive.org/web/20190501114826/http://www.igmchicago.org/surveys/robots-and-artificial-intelligence) from the original on 1 May 2019. Retrieved 3 July 2019.
- Lohr, Steve (2017). "Robots Will Take Jobs, but Not as Fast as Some Fear, New Report Says" (https://www.nytimes.com/2017/01/12/technology/robots-will-take-jobs-but-not-as-fast-as-some-fear-new-report-says.html). *The New York Times*. Archived (https://web.archive.org/web/20180114073704/https://www.nytimes.com/2017/01/12/technology/robots-will-take-jobs-but-not-as-fast-as-some-fear-new-report-says.html) from the original on 14 January 2018. Retrieved 13 January 2018.
- Frey, Carl Benedikt; Osborne, Michael A (1 January 2017). "The future of employment: How susceptible are jobs to computerisation?". *Technological Forecasting and Social Change*. **114**: 254–280. CiteSeerX 10.1.1.395.416 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.395.416). doi:10.1016/j.techfore.2016.08.019 (https://doi.org/10.1016%2Fj.techfore.2016.08.019). ISSN 0040-1625 (https://www.worldcat.org/issn/0040-1625).
- Arntz, Ulrich; Gregory; Zierahn (2016), "The risk of automation for jobs in OECD countries: A comparative analysis", *OECD Social, Employment, and Migration Working Papers 189*
- Morgenstern, Michael (9 May 2015). "Automation and anxiety" (https://www.economist.com/news/special-report/21700758-will-smarter-machines-cause-mass-unemployment-automation-and-anxiety). *The Economist*. Archived (https://web.archive.org/web/20180112214621/https://www.economist.com/news/special-report/21700758-will-smarter-machines-cause-mass-unemployment-automation-and-anxiety) from the original on 12 January 2018. Retrieved 13 January 2018.
- Mahdawi, Arwa (26 June 2017). "What jobs will still be around in 20 years? Read this to prepare your future" (https://www.theguardian.com/us-news/2017/jun/26/jobs-future-automation-robots-skills-creative-health). *The Guardian*. Archived (https://web.archive.org/web/20180114021804/https://www.theguardian.com/us-news/2017/jun/26/jobs-future-automation-robots-skills-creative-health) from the original on 14 January 2018. Retrieved 13 January 2018.
- Rubin, Charles (Spring 2003). "Artificial Intelligence and Human Nature" (https://web.archive.org/web/20120611115223/http://www.thenewatlantis.com/publications/artificial-intelligence-and-human-nature). *The New Atlantis*. **1**: 88–100. Archived from the original (http://www.thenewatlantis.com/publications/artificial-intelligence-and-human-nature) on 11 June 2012.
- Bostrom, Nick (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

- Brooks, Rodney (10 November 2014). "artificial intelligence is a tool, not a threat" (https://web.archive.org/web/20141112130954/http://www.rethinkrobotics.com/artificial-intelligence-tool-threat/). Archived from the original (http://www.rethinkrobotics.com/artificial-intelligence-tool-threat/) on 12 November 2014.
- Sainato, Michael (19 August 2015). "Stephen Hawking, Elon Musk, and Bill Gates Warn About Artificial Intelligence" (https://observer.com/2015/08/stephen-hawking-elon-musk-and-bill-gates-warn-about-artificial-intelligence/). *Observer*. Archived (https://web.archive.org/web/20151030053323/http://observer.com/2015/08/stephen-hawking-elon-musk-and-bill-gates-warn-about-artificial-intelligence/) from the original on 30 October 2015. Retrieved 30 October 2015.
- Harari, Yuval Noah (October 2018). "Why Technology Favors Tyranny" (https://www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technology-tyranny/568330/?utm_source=newsletter&utm_medium=email&utm_campaign=atlantic-daily-newsletter&utm_content=20180905&silverid-ref=MzEwMTU3MjI5NjU1S0). *The Atlantic*.
- Robitzski, Dan (5 September 2018). "Five experts share what scares them the most about AI" (https://futurism.com/artificial-intelligence-experts-fear/amp). Archived (https://web.archive.org/web/20191208094101/https://futurism.com/artificial-intelligence-experts-fear/amp) from the original on 8 December 2019. Retrieved 8 December 2019.
- Goffrey, Andrew (2008). "Algorithm". In Fuller, Matthew (ed.). *Software studies: a lexicon* (https://archive.org/details/softwarestudiesl00full_007). Cambridge, Mass.: MIT Press. pp. 15 (https://archive.org/details/softwarestudiesl00full_007/page/n29)–20. ISBN 978-1-4356-4787-9.
- Lipartito, Kenneth (6 January 2011). "The Narrative and the Algorithm: Genres of Credit Reporting from the Nineteenth Century to Today" (https://mpra.ub.uni-muenchen.de/28142/1/MPRA_paper_28142.pdf) (PDF) (Submitted manuscript). doi:10.2139/ssrn.1736283 (https://doi.org/10.2139%2Fssrn.1736283). S2CID 166742927 (https://api.semanticscholar.org/CorpusID:166742927).
- Goodman, Bryce; Flaxman, Seth (2017). "EU regulations on algorithmic decision-making and a "right to explanation" ". *AI Magazine*. **38** (3): 50. arXiv:1606.08813 (https://arxiv.org/abs/1606.08813). doi:10.1609/aimag.v38i3.2741 (https://doi.org/10.1609%2Faimag.v38i3.2741). S2CID 7373959 (https://api.semanticscholar.org/CorpusID:7373959).
- CNA (12 January 2019). "Commentary: Bad news. Artificial intelligence is biased" (https://www.channelnewsasia.com/news/commentary/artificial-intelligence-big-data-bias-hiring-loans-key-challenge-11097374). *CNA*. Archived (https://web.archive.org/web/20190112104421/https://www.channelnewsasia.com/news/commentary/artificial-intelligence-big-data-bias-hiring-loans-key-challenge-11097374) from the original on 12 January 2019. Retrieved 19 June 2020.
- Larson, Jeff; Angwin, Julia (23 May 2016). "How We Analyzed the COMPAS Recidivism Algorithm" (https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm). *ProPublica*. Archived (https://web.archive.org/web/20190429190950/https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm) from the original on 29 April 2019. Retrieved 19 June 2020.
- Müller, Vincent C.; Bostrom, Nick (2014). "Future Progress in Artificial Intelligence: A Poll Among Experts" (http://www.sophia.de/pdf/2014_PT-AI_polls.pdf) (PDF). *AI Matters*. **1** (1): 9–11. doi:10.1145/2639475.2639478 (https://doi.org/10.1145%2F2639475.2639478). S2CID 8510016 (https://api.semanticscholar.org/CorpusID:8510016). Archived (https://web.archive.org/web/20160115114604/http://www.sophia.de/pdf/2014_PT-AI_polls.pdf) (PDF) from the original on 15 January 2016.
- Cellan-Jones, Rory (2 December 2014). "Stephen Hawking warns artificial intelligence could end mankind" (https://www.bbc.com/news/technology-30290540). *BBC News*. Archived (https://web.archive.org/web/20151030054329/http://www.bbc.com/news/technology-30290540) from the original on 30 October 2015. Retrieved 30 October 2015.
- Rawlinson, Kevin (29 January 2015). "Microsoft's Bill Gates insists AI is a threat" (https://www.bbc.co.uk/news/31047780). *BBC News*. Archived (https://web.archive.org/web/201501291

83607/http://www.bbc.co.uk/news/31047780) from the original on 29 January 2015. Retrieved 30 January 2015.

- Holley, Peter (28 January 2015). "Bill Gates on dangers of artificial intelligence: 'I don't understand why some people are not concerned' " (https://www.washingtonpost.com/news/the-switch/wp/2015/01/28/bill-gates-on-dangers-of-artificial-intelligence-dont-understand-why-some-people-are-not-concerned/). *The Washington Post*. ISSN 0190-8286 (https://www.worldcat.org/issn/0190-8286). Archived (https://web.archive.org/web/20151030054330/https://www.washingtonpost.com/news/the-switch/wp/2015/01/28/bill-gates-on-dangers-of-artificial-intelligence-dont-understand-why-some-people-are-not-concerned/) from the original on 30 October 2015. Retrieved 30 October 2015.
- Gibbs, Samuel (27 October 2014). "Elon Musk: artificial intelligence is our biggest existential threat" (https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat). *The Guardian*. Archived (https://web.archive.org/web/20151030054330/http://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat) from the original on 30 October 2015. Retrieved 30 October 2015.
- Churm, Philip Andrew (14 May 2019). "Yuval Noah Harari talks politics, technology and migration" (https://www.euronews.com/2019/05/14/a-i-is-as-threatening-as-climate-change-and-nuclear-war-says-philosopher-yuval-noah-harari). *euronews*. Archived (https://web.archive.org/web/20190514192815/https://www.euronews.com/2019/05/14/a-i-is-as-threatening-as-climate-change-and-nuclear-war-says-philosopher-yuval-noah-harari) from the original on 14 May 2019. Retrieved 15 November 2020.
- Bostrom, Nick (2015). "What happens when our computers get smarter than we are?" (https://www.ted.com/talks/nick_bostrom_what_happens_when_our_computers_get_smarter_than_we_are/transcript). TED (conference). Archived (https://web.archive.org/web/20200725005719/https://www.ted.com/talks/nick_bostrom_what_happens_when_our_computers_get_smarter_than_we_are/transcript) from the original on 25 July 2020. Retrieved 30 January 2020.
- Post, Washington (2015). "Tech titans like Elon Musk are spending $1 billion to save you from terminators" (https://www.chicagotribune.com/bluesky/technology/ct-tech-titans-against-terminators-20151214-story.html). Archived (https://web.archive.org/web/20160607121118/http://www.chicagotribune.com/bluesky/technology/ct-tech-titans-against-terminators-20151214-story.html) from the original on 7 June 2016.
- Del Prado, Guia Marie (9 October 2015). "The mysterious artificial intelligence company Elon Musk invested in is developing game-changing smart computers" (http://www.techinsider.io/mysterious-artificial-intelligence-company-elon-musk-investment-2015-10). *Tech Insider*. Archived (https://web.archive.org/web/20151030165333/http://www.techinsider.io/mysterious-artificial-intelligence-company-elon-musk-investment-2015-10) from the original on 30 October 2015. Retrieved 30 October 2015.
- FastCompany (15 January 2015). "Elon Musk Is Donating $10M Of His Own Money To Artificial Intelligence Research" (http://www.fastcompany.com/3041007/fast-feed/elon-musk-is-donating-10m-of-his-own-money-to-artificial-intelligence-research). *Fast Company*. Archived (https://web.archive.org/web/20151030202356/http://www.fastcompany.com/3041007/fast-feed/elon-musk-is-donating-10m-of-his-own-money-to-artificial-intelligence-research) from the original on 30 October 2015. Retrieved 30 October 2015.
- Thibodeau, Patrick (25 March 2019). "Oracle CEO Mark Hurd sees no reason to fear ERP AI" (https://searcherp.techtarget.com/news/252460208/Oracle-CEO-Mark-Hurd-sees-no-reason-to-fear-ERP-AI). *SearchERP*. Archived (https://web.archive.org/web/20190506173749/https://searcherp.techtarget.com/news/252460208/Oracle-CEO-Mark-Hurd-sees-no-reason-to-fear-ERP-AI) from the original on 6 May 2019. Retrieved 6 May 2019.
- Bhardwaj, Prachi (24 May 2018). "Mark Zuckerberg responds to Elon Musk's paranoia about AI: 'AI is going to... help keep our communities safe.' " (https://www.businessinsider.com/mark-zuckerberg-shares-thoughts-elon-musks-ai-2018-5). *Business Insider*. Archived (https://we

b.archive.org/web/20190506173756/https://www.businessinsider.com/mark-zuckerberg-shares-thoughts-elon-musks-ai-2018-5) from the original on 6 May 2019. Retrieved 6 May 2019.

- Geist, Edward Moore (9 August 2015). "Is artificial intelligence really an existential threat to humanity?" (http://thebulletin.org/artificial-intelligence-really-existential-threat-humanity8577). *Bulletin of the Atomic Scientists*. Archived (https://web.archive.org/web/20151030054330/http://thebulletin.org/artificial-intelligence-really-existential-threat-humanity8577) from the original on 30 October 2015. Retrieved 30 October 2015.
- Madrigal, Alexis C. (27 February 2015). "The case against killer robots, from a guy actually working on artificial intelligence" (http://fusion.net/story/54583/the-case-against-killer-robots-from-a-guy-actually-building-ai/). *Fusion.net*. Archived (https://web.archive.org/web/20160204175716/http://fusion.net/story/54583/the-case-against-killer-robots-from-a-guy-actually-building-ai/) from the original on 4 February 2016. Retrieved 31 January 2016.
- Lee, Timothy B. (22 August 2014). "Will artificial intelligence destroy humanity? Here are 5 reasons not to worry" (https://www.vox.com/2014/8/22/6043635/5-reasons-we-shouldnt-worry-about-super-intelligent-computers-taking). *Vox*. Archived (https://web.archive.org/web/20151030092203/http://www.vox.com/2014/8/22/6043635/5-reasons-we-shouldnt-worry-about-super-intelligent-computers-taking) from the original on 30 October 2015. Retrieved 30 October 2015.
- Law Library of Congress (U.S.). Global Legal Research Directorate, issuing body. (2019). *Regulation of artificial intelligence in selected jurisdictions*. LCCN 2019668143 (https://lccn.loc.gov/2019668143). OCLC 1110727808 (https://www.worldcat.org/oclc/1110727808).
- *UNESCO Science Report: the Race Against Time for Smarter Development* (https://unesdoc.unesco.org/ark:/48223/pf0000377433/PDF/377433eng.pdf.multi). Paris: UNESCO. 11 June 2021. ISBN 978-92-3-100450-6.
- Berryhill, Jamie; Heang, Kévin Kok; Clogher, Rob; McBride, Keegan (2019). *Hello, World: Artificial Intelligence and its Use in the Public Sector* (https://oecd-opsi.org/wp-content/uploads/2019/11/AI-Report-Online.pdf) (PDF). Paris: OECD Observatory of Public Sector Innovation. Archived (https://web.archive.org/web/20191220021331/https://oecd-opsi.org/wp-content/uploads/2019/11/AI-Report-Online.pdf) (PDF) from the original on 20 December 2019. Retrieved 9 August 2020.
- Barfield, Woodrow; Pagallo, Ugo (2018). *Research handbook on the law of artificial intelligence*. Cheltenham, UK. ISBN 978-1-78643-904-8. OCLC 1039480085 (https://www.worldcat.org/oclc/1039480085).
- Iphofen, Ron; Kritikos, Mihalis (3 January 2019). "Regulating artificial intelligence and robotics: ethics by design in a digital society". *Contemporary Social Science*. **16** (2): 170–184. doi:10.1080/21582041.2018.1563803 (https://doi.org/10.1080%2F21582041.2018.1563803). ISSN 2158-2041 (https://www.worldcat.org/issn/2158-2041). S2CID 59298502 (https://api.semanticscholar.org/CorpusID:59298502).
- Wirtz, Bernd W.; Weyerer, Jan C.; Geyer, Carolin (24 July 2018). "Artificial Intelligence and the Public Sector – Applications and Challenges" (https://zenodo.org/record/3569435). *International Journal of Public Administration*. **42** (7): 596–615. doi:10.1080/01900692.2018.1498103 (https://doi.org/10.1080%2F01900692.2018.1498103). ISSN 0190-0692 (https://www.worldcat.org/issn/0190-0692). S2CID 158829602 (https://api.semanticscholar.org/CorpusID:158829602). Archived (https://web.archive.org/web/20200818131415/https://zenodo.org/record/3569435) from the original on 18 August 2020. Retrieved 22 August 2020.
- Buiten, Miriam C (2019). "Towards Intelligent Regulation of Artificial Intelligence" (https://doi.org/10.1017%2Ferr.2019.8). *European Journal of Risk Regulation*. **10** (1): 41–59. doi:10.1017/err.2019.8 (https://doi.org/10.1017%2Ferr.2019.8). ISSN 1867-299X (https://www.worldcat.org/issn/1867-299X).
- Wallach, Wendell (2010). *Moral Machines*. Oxford University Press.
- Brown, Eileen (5 November 2019). "Half of Americans do not believe deepfake news could target them online" (https://www.zdnet.com/article/half-of-americans-do-not-believe-deepfak

e-news-could-target-them-online/). *ZDNet*. Archived (https://web.archive.org/web/20191106035012/https://www.zdnet.com/article/half-of-americans-do-not-believe-deepfake-news-could-target-them-online/) from the original on 6 November 2019. Retrieved 3 December 2019.

- Frangoul, Anmar (14 June 2019). "A Californian business is using A.I. to change the way we think about energy storage" (https://www.cnbc.com/2019/06/14/the-business-using-ai-to-change-how-we-think-about-energy-storage.html). *CNBC*. Archived (https://web.archive.org/web/20200725044735/https://www.cnbc.com/2019/06/14/the-business-using-ai-to-change-how-we-think-about-energy-storage.html) from the original on 25 July 2020. Retrieved 5 November 2019.
- "The Economist Explains: Why firms are piling into artificial intelligence" (https://www.economist.com/blogs/economist-explains/2016/04/economist-explains). *The Economist*. 31 March 2016. Archived (https://web.archive.org/web/20160508010311/http://www.economist.com/blogs/economist-explains/2016/04/economist-explains) from the original on 8 May 2016. Retrieved 19 May 2016.
- Lohr, Steve (28 February 2016). "The Promise of Artificial Intelligence Unfolds in Small Steps" (https://www.nytimes.com/2016/02/29/technology/the-promise-of-artificial-intelligence-unfolds-in-small-steps.html). *The New York Times*. Archived (https://web.archive.org/web/20160229171843/http://www.nytimes.com/2016/02/29/technology/the-promise-of-artificial-intelligence-unfolds-in-small-steps.html) from the original on 29 February 2016. Retrieved 29 February 2016.
- Smith, Mark (22 July 2016). "So you think you chose to read this article?" (https://www.bbc.co.uk/news/business-36837824). *BBC News*. Archived (https://web.archive.org/web/20160725205007/http://www.bbc.co.uk/news/business-36837824) from the original on 25 July 2016.
- Aletras, N.; Tsarapatsanis, D.; Preotiuc-Pietro, D.; Lampos, V. (2016). "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective" (https://doi.org/10.7717%2Fpeerj-cs.93). *PeerJ Computer Science*. **2**: e93. doi:10.7717/peerj-cs.93 (https://doi.org/10.7717%2Fpeerj-cs.93).
- Cadena, Cesar; Carlone, Luca; Carrillo, Henry; Latif, Yasir; Scaramuzza, Davide; Neira, Jose; Reid, Ian; Leonard, John J. (December 2016). "Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age". *IEEE Transactions on Robotics*. **32** (6): 1309–1332. arXiv:1606.05830 (https://arxiv.org/abs/1606.05830). Bibcode:2016arXiv160605830C (https://ui.adsabs.harvard.edu/abs/2016arXiv160605830C). doi:10.1109/TRO.2016.2624754 (https://doi.org/10.1109%2FTRO.2016.2624754). S2CID 2596787 (https://api.semanticscholar.org/CorpusID:2596787).
- Cambria, Erik; White, Bebo (May 2014). "Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]". *IEEE Computational Intelligence Magazine*. **9** (2): 48–57. doi:10.1109/MCI.2014.2307227 (https://doi.org/10.1109%2FMCI.2014.2307227). S2CID 206451986 (https://api.semanticscholar.org/CorpusID:206451986).
- Vincent, James (7 November 2019). "OpenAI has published the text-generating AI it said was too dangerous to share" (https://www.theverge.com/2019/11/7/20953040/openai-text-generation-ai-gpt-2-full-model-release-1-5b-parameters). *The Verge*. Archived (https://web.archive.org/web/20200611054114/https://www.theverge.com/2019/11/7/20953040/openai-text-generation-ai-gpt-2-full-model-release-1-5b-parameters) from the original on 11 June 2020. Retrieved 11 June 2020.
- Jordan, M. I.; Mitchell, T. M. (16 July 2015). "Machine learning: Trends, perspectives, and prospects". *Science*. **349** (6245): 255–260. Bibcode:2015Sci...349..255J (https://ui.adsabs.harvard.edu/abs/2015Sci...349..255J). doi:10.1126/science.aaa8415 (https://doi.org/10.1126%2Fscience.aaa8415). PMID 26185243 (https://pubmed.ncbi.nlm.nih.gov/26185243). S2CID 677218 (https://api.semanticscholar.org/CorpusID:677218).
- Maschafilm (2010). "Content: Plug & Pray Film – Artificial Intelligence – Robots -" (http://www.plugandpray-film.de/en/content.html). *plugandpray-film.de*. Archived (https://web.archive.org/web/20160212040134/http://www.plugandpray-film.de/en/content.html) from the original on 12 February 2016.

- Evans, Woody (2015). "Posthuman Rights: Dimensions of Transhuman Worlds" (https://doi.org/10.5209%2Frev_TK.2015.v12.n2.49072). *Teknokultura*. **12** (2). doi:10.5209/rev_TK.2015.v12.n2.49072 (https://doi.org/10.5209%2Frev_TK.2015.v12.n2.49072).
- Waddell, Kaveh (2018). "Chatbots Have Entered the Uncanny Valley" (https://www.theatlantic.com/technology/archive/2017/04/uncanny-valley-digital-assistants/523806/). *The Atlantic*. Archived (https://web.archive.org/web/20180424202350/https://www.theatlantic.com/technology/archive/2017/04/uncanny-valley-digital-assistants/523806/) from the original on 24 April 2018. Retrieved 24 April 2018.
- Poria, Soujanya; Cambria, Erik; Bajpai, Rajiv; Hussain, Amir (September 2017). "A review of affective computing: From unimodal analysis to multimodal fusion" (http://researchrepository.napier.ac.uk/Output/1792429). *Information Fusion*. **37**: 98–125. doi:10.1016/j.inffus.2017.02.003 (https://doi.org/10.1016%2Fj.inffus.2017.02.003). hdl:1893/25490 (https://hdl.handle.net/1893%2F25490).
- "Robots could demand legal rights" (http://news.bbc.co.uk/2/hi/technology/6200005.stm). *BBC News*. 21 December 2006. Archived (https://web.archive.org/web/20191015042628/http://news.bbc.co.uk/2/hi/technology/6200005.stm) from the original on 15 October 2019. Retrieved 3 February 2011.
- Horst, Steven (2005). "The Computational Theory of Mind" (http://plato.stanford.edu/entries/computational-mind). *The Stanford Encyclopedia of Philosophy*.
- Omohundro, Steve (2008). *The Nature of Self-Improving Artificial Intelligence*. presented and distributed at the 2007 Singularity Summit, San Francisco, CA.
- Ford, Martin; Colvin, Geoff (6 September 2015). "Will robots create more jobs than they destroy?" (https://www.theguardian.com/technology/2015/sep/06/will-robots-create-destroy-jobs). *The Guardian*. Archived (https://web.archive.org/web/20180616204119/https://www.theguardian.com/technology/2015/sep/06/will-robots-create-destroy-jobs) from the original on 16 June 2018. Retrieved 13 January 2018.
- *White Paper: On Artificial Intelligence – A European approach to excellence and trust* (https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf) (PDF). Brussels: European Commission. 2020. Archived (https://web.archive.org/web/20200220173419/https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf) (PDF) from the original on 20 February 2020. Retrieved 20 February 2020.
- Anderson, Michael; Anderson, Susan Leigh (2011). *Machine Ethics*. Cambridge University Press.
- "Machine Ethics" (https://web.archive.org/web/20141129044821/http://www.aaai.org/Library/Symposia/Fall/fs05-06). *aaai.org*. Archived from the original (http://www.aaai.org/Library/Symposia/Fall/fs05-06) on 29 November 2014.
- Russell, Stuart (8 October 2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. United States: Viking. ISBN 978-0-525-55861-3. OCLC 1083694322 (https://www.worldcat.org/oclc/1083694322).
- "AI set to exceed human brain power" (http://www.cnn.com/2006/TECH/science/07/24/ai.bostrom/). *CNN*. 9 August 2006. Archived (https://web.archive.org/web/20080219001624/http://www.cnn.com/2006/TECH/science/07/24/ai.bostrom/) from the original on 19 February 2008.
- "Robots could demand legal rights" (http://news.bbc.co.uk/2/hi/technology/6200005.stm). *BBC News*. 21 December 2006. Archived (https://web.archive.org/web/20191015042628/http://news.bbc.co.uk/2/hi/technology/6200005.stm) from the original on 15 October 2019. Retrieved 3 February 2011.
- "Kismet" (http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html). MIT Artificial Intelligence Laboratory, Humanoid Robotics Group. Archived (https://web.archive.or

g/web/20141017040432/http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html) from the original on 17 October 2014. Retrieved 25 October 2014.

- Sikos, Leslie F. (June 2017). *Description Logics in Multimedia Reasoning* (https://www.springer.com/us/book/9783319540658). Cham: Springer. doi:10.1007/978-3-319-54066-5 (https://doi.org/10.1007%2F978-3-319-54066-5). ISBN 978-3-319-54066-5. S2CID 3180114 (https://api.semanticscholar.org/CorpusID:3180114). Archived (https://web.archive.org/web/20170829120912/https://www.springer.com/us/book/9783319540658) from the original on 29 August 2017.
- Smoliar, Stephen W.; Zhang, HongJiang (1994). "Content based video indexing and retrieval". *IEEE Multimedia*. **1** (2): 62–72. doi:10.1109/93.311653 (https://doi.org/10.1109%2F93.311653). S2CID 32710913 (https://api.semanticscholar.org/CorpusID:32710913).
- Neumann, Bernd; Möller, Ralf (January 2008). "On scene interpretation with description logics". *Image and Vision Computing*. **26** (1): 82–101. doi:10.1016/j.imavis.2007.08.013 (https://doi.org/10.1016%2Fj.imavis.2007.08.013).
- Kuperman, G. J.; Reichley, R. M.; Bailey, T. C. (1 July 2006). "Using Commercial Knowledge Bases for Clinical Decision Support: Opportunities, Hurdles, and Recommendations" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1513681). *Journal of the American Medical Informatics Association*. **13** (4): 369–371. doi:10.1197/jamia.M2055 (https://doi.org/10.1197%2Fjamia.M2055). PMC 1513681 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1513681). PMID 16622160 (https://pubmed.ncbi.nlm.nih.gov/16622160).
- McGarry, Ken (1 December 2005). "A survey of interestingness measures for knowledge discovery". *The Knowledge Engineering Review*. **20** (1): 39–61. doi:10.1017/S0269888905000408 (https://doi.org/10.1017%2FS0269888905000408). S2CID 14987656 (https://api.semanticscholar.org/CorpusID:14987656).
- Bertini, M; Del Bimbo, A; Torniai, C (2006). "Automatic annotation and semantic retrieval of video sequences using multimedia ontologies". *MM '06 Proceedings of the 14th ACM international conference on Multimedia*. 14th ACM international conference on Multimedia. Santa Barbara: ACM. pp. 679–682.
- Kahneman, Daniel (25 October 2011). *Thinking, Fast and Slow* (https://books.google.com/books?id=ZuKTvERuPG8C). Macmillan. ISBN 978-1-4299-6935-2. Retrieved 8 April 2012.
- Turing, Alan (1948), "Machine Intelligence", in Copeland, B. Jack (ed.), *The Essential Turing: The ideas that gave birth to the computer age*, Oxford: Oxford University Press, p. 412, ISBN 978-0-19-825080-7
- Domingos, Pedro (22 September 2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books. ISBN 978-0465065707.
- Minsky, Marvin (1986), *The Society of Mind*, Simon and Schuster
- Pinker, Steven (4 September 2007) [1994], *The Language Instinct*, Perennial Modern Classics, Harper, ISBN 978-0-06-133646-1
- Chalmers, David (1995). "Facing up to the problem of consciousness" (http://www.imprint.co.uk/chalmers.html). *Journal of Consciousness Studies*. **2** (3): 200–219. Archived (https://web.archive.org/web/20050308163649/http://www.imprint.co.uk/chalmers.html) from the original on 8 March 2005. Retrieved 11 October 2018.
- Roberts, Jacob (2016). "Thinking Machines: The Search for Artificial Intelligence" (https://web.archive.org/web/20180819152455/https://www.sciencehistory.org/distillations/magazine/thinking-machines-the-search-for-artificial-intelligence). *Distillations*. Vol. 2 no. 2. pp. 14–23. Archived from the original (https://www.sciencehistory.org/distillations/magazine/thinking-machines-the-search-for-artificial-intelligence) on 19 August 2018. Retrieved 20 March 2018.
- Pennachin, C.; Goertzel, B. (2007). "Contemporary Approaches to Artificial General Intelligence". *Artificial General Intelligence*. Cognitive Technologies. Berlin, Heidelberg: Springer. doi:10.1007/978-3-540-68677-4_1 (https://doi.org/10.1007%2F978-3-540-68677-4_1). ISBN 978-3-540-23733-4.

- "Ask the AI experts: What's driving today's progress in AI?" (https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/ask-the-ai-experts-whats-driving-todays-progress-in-ai). *McKinsey & Company*. Archived (https://web.archive.org/web/20180413190018/https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/ask-the-ai-experts-whats-driving-todays-progress-in-ai) from the original on 13 April 2018. Retrieved 13 April 2018.
- "Reshaping Business With Artificial Intelligence" (https://sloanreview.mit.edu/projects/reshaping-business-with-artificial-intelligence/). *MIT Sloan Management Review*. Archived (https://web.archive.org/web/20180519171905/https://sloanreview.mit.edu/projects/reshaping-business-with-artificial-intelligence/) from the original on 19 May 2018. Retrieved 2 May 2018.
- Lorica, Ben (18 December 2017). "The state of AI adoption" (https://www.oreilly.com/ideas/the-state-of-ai-adoption). *O'Reilly Media*. Archived (https://web.archive.org/web/20180502140700/https://www.oreilly.com/ideas/the-state-of-ai-adoption) from the original on 2 May 2018. Retrieved 2 May 2018.
- "AlphaGo – Google DeepMind" (https://deepmind.com/alpha-go.html). Archived (https://web.archive.org/web/20160310191926/https://www.deepmind.com/alpha-go.html) from the original on 10 March 2016.

- Asada, M.; Hosoda, K.; Kuniyoshi, Y.; Ishiguro, H.; Inui, T.; Yoshikawa, Y.; Ogino, M.; Yoshida, C. (2009). "Cognitive developmental robotics: a survey". *IEEE Transactions on Autonomous Mental Development*. **1** (1): 12–34. doi:10.1109/tamd.2009.2021702 (https://doi.org/10.1109%2Ftamd.2009.2021702). S2CID 10168773 (https://api.semanticscholar.org/CorpusID:10168773).
- Ashok83 (10 September 2019). "How AI Is Getting Groundbreaking Changes In Talent Management And HR Tech" (https://hackernoon.com/how-ai-is-getting-groundbreaking-changes-in-talent-management-and-hr-tech-d24ty3zzd). Hackernoon. Archived (https://web.archive.org/web/20190911083500/https://hackernoon.com/how-ai-is-getting-groundbreaking-changes-in-talent-management-and-hr-tech-d24ty3zzd) from the original on 11 September 2019. Retrieved 14 February 2020.
- Berlinski, David (2000). *The Advent of the Algorithm* (https://archive.org/details/adventofalgorith0000berl). Harcourt Books. ISBN 978-0-15-601391-8. OCLC 46890682 (https://www.worldcat.org/oclc/46890682). Archived (https://web.archive.org/web/20200726215744/https://archive.org/details/adventofalgorith0000berl) from the original on 26 July 2020. Retrieved 22 August 2020.
- Brooks, Rodney (1990). "Elephants Don't Play Chess" (http://people.csail.mit.edu/brooks/papers/elephants.pdf) (PDF). *Robotics and Autonomous Systems*. **6** (1–2): 3–15. CiteSeerX 10.1.1.588.7539 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.588.7539). doi:10.1016/S0921-8890(05)80025-9 (https://doi.org/10.1016%2FS0921-8890%2805%2980025-9). Archived (https://web.archive.org/web/20070809020912/http://people.csail.mit.edu/brooks/papers/elephants.pdf) (PDF) from the original on 9 August 2007.
- Butler, Samuel (13 June 1863). "Darwin among the Machines" (http://www.nzetc.org/tm/scholarly/tei-ButFir-t1-g1-t1-g1-t4-body.html). Letters to the Editor. *The Press*. Christchurch, New Zealand. Archived (https://web.archive.org/web/20080919172551/http://www.nzetc.org/tm/scholarly/tei-ButFir-t1-g1-t1-g1-t4-body.html) from the original on 19 September 2008. Retrieved 16 October 2014 – via Victoria University of Wellington.
- Clark, Jack (1 July 2015a). "Musk-Backed Group Probes Risks Behind Artificial Intelligence" (https://www.bloomberg.com/news/articles/2015-07-01/musk-backed-group-probes-risks-behind-artificial-intelligence). *Bloomberg.com*. Archived (https://web.archive.org/web/20151030202356/http://www.bloomberg.com/news/articles/2015-07-01/musk-backed-group-probes-risks-behind-artificial-intelligence) from the original on 30 October 2015. Retrieved 30 October 2015.
- Clark, Jack (8 December 2015b). "Why 2015 Was a Breakthrough Year in Artificial Intelligence" (https://www.bloomberg.com/news/articles/2015-12-08/why-2015-was-a-breakt

hrough-year-in-artificial-intelligence). *Bloomberg.com*. Archived (https://web.archive.org/web/20161123053855/https://www.bloomberg.com/news/articles/2015-12-08/why-2015-was-a-breakthrough-year-in-artificial-intelligence) from the original on 23 November 2016. Retrieved 23 November 2016.

- Dennett, Daniel (1991). *Consciousness Explained*. The Penguin Press. ISBN 978-0-7139-9037-9.
- Dreyfus, Hubert (1972). *What Computers Can't Do*. New York: MIT Press. ISBN 978-0-06-011082-6.
- Dreyfus, Hubert; Dreyfus, Stuart (1986). *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer* (https://archive.org/details/mindovermachinep00drey). Oxford, UK: Blackwell. ISBN 978-0-02-908060-3. Archived (https://web.archive.org/web/20200726131414/https://archive.org/details/mindovermachinep00drey) from the original on 26 July 2020. Retrieved 22 August 2020.
- Dyson, George (1998). *Darwin among the Machines* (https://archive.org/details/darwinamongmachi00dyso). Allan Lane Science. ISBN 978-0-7382-0030-9. Archived (https://web.archive.org/web/20200726131443/https://archive.org/details/darwinamongmachi00dyso) from the original on 26 July 2020. Retrieved 22 August 2020.
- Edelson, Edward (1991). *The Nervous System* (https://archive.org/details/nervoussystem0000edel). New York: Chelsea House. ISBN 978-0-7910-0464-7. Archived (https://web.archive.org/web/20200726131758/https://archive.org/details/nervoussystem0000edel) from the original on 26 July 2020. Retrieved 18 November 2019.
- Fearn, Nicholas (2007). *The Latest Answers to the Oldest Questions: A Philosophical Adventure with the World's Greatest Thinkers*. New York: Grove Press. ISBN 978-0-8021-1839-4.
- Haugeland, John (1985). *Artificial Intelligence: The Very Idea*. Cambridge, Mass.: MIT Press. ISBN 978-0-262-08153-5.
- Hawkins, Jeff; Blakeslee, Sandra (2005). *On Intelligence*. New York: Owl Books. ISBN 978-0-8050-7853-4.
- Henderson, Mark (24 April 2007). "Human rights for robots? We're getting carried away" (http://www.thetimes.co.uk/tto/technology/article1966391.ece). *The Times Online*. London. Archived (https://web.archive.org/web/20140531104850/http://www.thetimes.co.uk/tto/technology/article1966391.ece) from the original on 31 May 2014. Retrieved 31 May 2014.
- Kahneman, Daniel; Slovic, D.; Tversky, Amos (1982). *Judgment under uncertainty: Heuristics and biases*. *Science*. **185**. New York: Cambridge University Press. pp. 1124–1131. doi:10.1126/science.185.4157.1124 (https://doi.org/10.1126%2Fscience.185.4157.1124). ISBN 978-0-521-28414-1. PMID 17835457 (https://pubmed.ncbi.nlm.nih.gov/17835457). S2CID 143452957 (https://api.semanticscholar.org/CorpusID:143452957).
- Katz, Yarden (1 November 2012). "Noam Chomsky on Where Artificial Intelligence Went Wrong" (https://www.theatlantic.com/technology/archive/2012/11/noam-chomsky-on-where-artificial-intelligence-went-wrong/261637/?single_page=true). *The Atlantic*. Archived (https://web.archive.org/web/20190228154403/https://www.theatlantic.com/technology/archive/2012/11/noam-chomsky-on-where-artificial-intelligence-went-wrong/261637/?single_page=true) from the original on 28 February 2019. Retrieved 26 October 2014.
- Kurzweil, Ray (2005). *The Singularity is Near*. Penguin Books. ISBN 978-0-670-03384-3.
- Langley, Pat (2011). "The changing science of machine learning" (https://doi.org/10.1007%2Fs10994-011-5242-y). *Machine Learning*. **82** (3): 275–279. doi:10.1007/s10994-011-5242-y (https://doi.org/10.1007%2Fs10994-011-5242-y).
- Legg, Shane; Hutter, Marcus (15 June 2007). *A Collection of Definitions of Intelligence* (Technical report). IDSIA. arXiv:0706.3639 (https://arxiv.org/abs/0706.3639). Bibcode:2007arXiv0706.3639L (https://ui.adsabs.harvard.edu/abs/2007arXiv0706.3639L). 07-07.

- Lenat, Douglas; Guha, R. V. (1989). *Building Large Knowledge-Based Systems*. Addison-Wesley. ISBN 978-0-201-51752-1.
- Lighthill, James (1973). "Artificial Intelligence: A General Survey". *Artificial Intelligence: a paper symposium*. Science Research Council.
- Lombardo, P; Boehm, I; Nairz, K (2020). "RadioComics – Santa Claus and the future of radiology" (https://doi.org/10.1016%2Fj.ejrad.2019.108771). *Eur J Radiol*. **122** (1): 108771. doi:10.1016/j.ejrad.2019.108771 (https://doi.org/10.1016%2Fj.ejrad.2019.108771). PMID 31835078 (https://pubmed.ncbi.nlm.nih.gov/31835078).
- Lungarella, M.; Metta, G.; Pfeifer, R.; Sandini, G. (2003). "Developmental robotics: a survey". *Connection Science*. **15** (4): 151–190. CiteSeerX 10.1.1.83.7615 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.83.7615). doi:10.1080/09540090310001655110 (https://doi.org/10.1080%2F09540090310001655110). S2CID 1452734 (https://api.semanticscholar.org/CorpusID:1452734).
- Maker, Meg Houston (2006). "AI@50: AI Past, Present, Future" (https://web.archive.org/web/20070103222615/http://www.engagingexperience.com/2006/07/ai50_ai_past_pr.html). Dartmouth College. Archived from the original (http://www.engagingexperience.com/2006/07/ai50_ai_past_pr.html) on 3 January 2007. Retrieved 16 October 2008.
- McCarthy, John; Minsky, Marvin; Rochester, Nathan; Shannon, Claude (1955). "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence" (https://web.archive.org/web/20070826230310/http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html). Archived from the original (http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html) on 26 August 2007. Retrieved 30 August 2007.
- Minsky, Marvin (1967). *Computation: Finite and Infinite Machines* (https://archive.org/details/computationfinit0000mins). Englewood Cliffs, N.J.: Prentice-Hall. ISBN 978-0-13-165449-5. Archived (https://web.archive.org/web/20200726131743/https://archive.org/details/computationfinit0000mins) from the original on 26 July 2020. Retrieved 18 November 2019.
- Moravec, Hans (1988). *Mind Children* (https://archive.org/details/mindchildrenfutu00mora). Harvard University Press. ISBN 978-0-674-57616-2. Archived (https://web.archive.org/web/20200726131644/https://archive.org/details/mindchildrenfutu00mora) from the original on 26 July 2020. Retrieved 18 November 2019.
- NRC (United States National Research Council) (1999). "Developments in Artificial Intelligence". *Funding a Revolution: Government Support for Computing Research*. National Academy Press.
- Newell, Allen; Simon, H. A. (1976). "Computer Science as Empirical Inquiry: Symbols and Search" (https://doi.org/10.1145%2F360018.360022). *Communications of the ACM*. **19** (3): 113–126. doi:10.1145/360018.360022 (https://doi.org/10.1145%2F360018.360022)..
- Nilsson, Nils (1983). "Artificial Intelligence Prepares for 2001" (https://ai.stanford.edu/~nilsson/OnlinePubs-Nils/General%20Essays/AIMag04-04-002.pdf) (PDF). *AI Magazine*. **1** (1). Archived (https://web.archive.org/web/20200817194457/http://ai.stanford.edu/~nilsson/OnlinePubs-Nils/General%20Essays/AIMag04-04-002.pdf) (PDF) from the original on 17 August 2020. Retrieved 22 August 2020. Presidential Address to the Association for the Advancement of Artificial Intelligence.
- Oudeyer, P-Y. (2010). "On the impact of robotics in behavioral and cognitive sciences: from insect navigation to human cognitive development" (http://www.pyoudeyer.com/IEEETAMDOudeyer10.pdf) (PDF). *IEEE Transactions on Autonomous Mental Development*. **2** (1): 2–16. doi:10.1109/tamd.2009.2039057 (https://doi.org/10.1109%2Ftamd.2009.2039057). S2CID 6362217 (https://api.semanticscholar.org/CorpusID:6362217). Archived (https://web.archive.org/web/20181003202543/http://www.pyoudeyer.com/IEEETAMDOudeyer10.pdf) (PDF) from the original on 3 October 2018. Retrieved 4 June 2013.
- Schank, Roger C. (1991). "Where's the AI". *AI magazine*. Vol. 12 no. 4.
- Searle, John (1980). "Minds, Brains and Programs" (http://cogprints.org/7150/1/10.1.1.83.5248.pdf) (PDF). *Behavioral and Brain Sciences*. **3** (3): 417–457.

doi:10.1017/S0140525X00005756 (https://doi.org/10.1017%2FS0140525X00005756). Archived (https://web.archive.org/web/20190317230215/http://cogprints.org/7150/1/10.1.1.83.5248.pdf) (PDF) from the original on 17 March 2019. Retrieved 22 August 2020.

- Searle, John (1999). *Mind, language and society* (https://archive.org/details/mindlanguagesoci00sear). New York: Basic Books. ISBN 978-0-465-04521-1. OCLC 231867665 (https://www.worldcat.org/oclc/231867665). Archived (https://web.archive.org/web/20200726220615/https://archive.org/details/mindlanguagesoci00sear) from the original on 26 July 2020. Retrieved 22 August 2020.
- Simon, H. A. (1965). *The Shape of Automation for Men and Management* (https://archive.org/details/shapeofautomatio00simo). New York: Harper & Row. Archived (https://web.archive.org/web/20200726131655/https://archive.org/details/shapeofautomatio00simo) from the original on 26 July 2020. Retrieved 18 November 2019.
- Solomonoff, Ray (1956). *An Inductive Inference Machine* (http://world.std.com/~rjs/indinf56.pdf) (PDF). Dartmouth Summer Research Conference on Artificial Intelligence. Archived (https://web.archive.org/web/20110426161749/http://world.std.com/~rjs/indinf56.pdf) (PDF) from the original on 26 April 2011. Retrieved 22 March 2011 – via std.com, pdf scanned copy of the original. Later published as
Solomonoff, Ray (1957). "An Inductive Inference Machine". *IRE Convention Record*. Section on Information Theory, part 2. pp. 56–62.
- Spadafora, Anthony (21 October 2016). "Stephen Hawking believes AI could be mankind's last accomplishment" (https://betanews.com/2016/10/21/artificial-intelligence-stephen-hawking/). *BetaNews*. Archived (https://web.archive.org/web/20170828183930/https://betanews.com/2016/10/21/artificial-intelligence-stephen-hawking/) from the original on 28 August 2017.
- Tao, Jianhua; Tan, Tieniu (2005). *Affective Computing and Intelligent Interaction*. Affective Computing: A Review. LNCS 3784. Springer. pp. 981–995. doi:10.1007/11573548 (https://doi.org/10.1007%2F11573548).
- Tecuci, Gheorghe (March–April 2012). "Artificial Intelligence". *Wiley Interdisciplinary Reviews: Computational Statistics*. **4** (2): 168–180. doi:10.1002/wics.200 (https://doi.org/10.1002%2Fwics.200).
- Thro, Ellen (1993). *Robotics: The Marriage of Computers and Machines* (https://archive.org/details/isbn_9780816026289). New York: Facts on File. ISBN 978-0-8160-2628-9. Archived (https://web.archive.org/web/20200726131505/https://archive.org/details/isbn_9780816026289) from the original on 26 July 2020. Retrieved 22 August 2020.
- Turing, Alan (October 1950), "Computing Machinery and Intelligence", *Mind*, **LIX** (236): 433–460, doi:10.1093/mind/LIX.236.433 (https://doi.org/10.1093%2Fmind%2FLIX.236.433), ISSN 0026-4423 (https://www.worldcat.org/issn/0026-4423).
- Vinge, Vernor (1993). "The Coming Technological Singularity: How to Survive in the Post-Human Era" (https://web.archive.org/web/20070101133646/http://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.html). *Vision 21: Interdisciplinary Science and Engineering in the Era of Cyberspace*: 11. Bibcode:1993vise.nasa...11V (https://ui.adsabs.harvard.edu/abs/1993vise.nasa...11V). Archived from the original (http://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.html) on 1 January 2007. Retrieved 14 November 2011.
- Wason, P. C.; Shapiro, D. (1966). "Reasoning" (https://archive.org/details/newhorizonsinpsy0000foss). In Foss, B. M. (ed.). *New horizons in psychology*. Harmondsworth: Penguin. Archived (https://web.archive.org/web/20200726131518/https://archive.org/details/newhorizonsinpsy0000foss) from the original on 26 July 2020. Retrieved 18 November 2019.
- Weng, J.; McClelland; Pentland, A.; Sporns, O.; Stockman, I.; Sur, M.; Thelen, E. (2001). "Autonomous mental development by robots and animals" (http://www.cse.msu.edu/dl/SciencePaper.pdf) (PDF). *Science*. **291** (5504): 599–600. doi:10.1126/science.291.5504.599 (https://doi.org/10.1126%2Fscience.291.5504.599). PMID 11229402 (https://pubmed.ncbi.nlm.nih.gov/11229402). S2CID 54131797 (https://api.semanticscholar.org/CorpusID:54131797). Archived (https://web.archive.org/web/20130904235242/http://www.cse.msu.edu/dl/Science

Paper.pdf) (PDF) from the original on 4 September 2013. Retrieved 4 June 2013 – via msu.edu.

Further reading

- DH Author, "Why Are There Still So Many Jobs? The History and Future of Workplace Automation" (2015) 29(3) *Journal of Economic Perspectives* 3.
- Boden, Margaret, *Mind As Machine*, Oxford University Press, 2006.
- Cukier, Kenneth, "Ready for Robots? How to Think about the Future of AI", *Foreign Affairs*, vol. 98, no. 4 (July/August 2019), pp. 192–98. George Dyson, historian of computing, writes (in what might be called "Dyson's Law") that "Any system simple enough to be understandable will not be complicated enough to behave intelligently, while any system complicated enough to behave intelligently will be too complicated to understand." (p. 197.) Computer scientist Alex Pentland writes: "Current AI machine-learning algorithms are, at their core, dead simple stupid. They work, but they work by brute force." (p. 198.)
- Domingos, Pedro, "Our Digital Doubles: AI will serve our species, not control it", *Scientific American*, vol. 319, no. 3 (September 2018), pp. 88–93.
- Gopnik, Alison, "Making AI More Human: Artificial intelligence has staged a revival by starting to incorporate what we know about how children learn", *Scientific American*, vol. 316, no. 6 (June 2017), pp. 60–65.
- Johnston, John (2008) *The Allure of Machinic Life: Cybernetics, Artificial Life, and the New AI*, MIT Press.
- Koch, Christof, "Proust among the Machines", *Scientific American*, vol. 321, no. 6 (December 2019), pp. 46–49. Christof Koch doubts the possibility of "intelligent" machines attaining consciousness, because "[e]ven the most sophisticated brain simulations are unlikely to produce conscious feelings." (p. 48.) According to Koch, "Whether machines can become sentient [is important] for ethical reasons. If computers experience life through their own senses, they cease to be purely a means to an end determined by their usefulness to... humans. Per GNW [the Global Neuronal Workspace theory], they turn from mere objects into subjects... with a point of view.... Once computers' cognitive abilities rival those of humanity, their impulse to push for legal and political rights will become irresistible—the right not to be deleted, not to have their memories wiped clean, not to suffer pain and degradation. The alternative, embodied by IIT [Integrated Information Theory], is that computers will remain only supersophisticated machinery, ghostlike empty shells, devoid of what we value most: the feeling of life itself." (p. 49.)
- Marcus, Gary, "Am I Human?: Researchers need new ways to distinguish artificial intelligence from the natural kind", *Scientific American*, vol. 316, no. 3 (March 2017), pp. 58–63. A stumbling block to AI has been an incapacity for reliable disambiguation. An example is the "pronoun disambiguation problem": a machine has no way of determining to whom or what a pronoun in a sentence refers. (p. 61.)
- E McGaughey, 'Will Robots Automate Your Job Away? Full Employment, Basic Income, and Economic Democracy' (2018) SSRN, part 2(3) (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3044448) Archived (https://web.archive.org/web/20180524201340/https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3044448) 24 May 2018 at the Wayback Machine.
- George Musser, "Artificial Imagination: How machines could learn creativity and common sense, among other human qualities", *Scientific American*, vol. 320, no. 5 (May 2019), pp. 58–63.
- Myers, Courtney Boyd ed. (2009). "The AI Report" (https://www.forbes.com/2009/06/22/singularity-robots-computers-opinions-contributors-artificial-intelligence-09_land.html) Archived (https://web.archive.org/web/20170729114303/https://www.forbes.com/2009/06/22/singularity-robots-computers-opinions-contributors-artificial-intelligence-09_land.html) 29 July 2017 at the Wayback Machine. *Forbes* June 2009

- Raphael, Bertram (1976). *The Thinking Computer* (https://archive.org/details/thinkingcomputer00raph). W.H. Freeman and Co. ISBN 978-0716707233. Archived (https://web.archive.org/web/20200726215746/https://archive.org/details/thinkingcomputer00raph) from the original on 26 July 2020. Retrieved 22 August 2020.
- Scharre, Paul, "Killer Apps: The Real Dangers of an AI Arms Race", *Foreign Affairs*, vol. 98, no. 3 (May/June 2019), pp. 135–44. "Today's AI technologies are powerful but unreliable. Rules-based systems cannot deal with circumstances their programmers did not anticipate. Learning systems are limited by the data on which they were trained. AI failures have already led to tragedy. Advanced autopilot features in cars, although they perform well in some circumstances, have driven cars without warning into trucks, concrete barriers, and parked cars. In the wrong situation, AI systems go from supersmart to superdumb in an instant. When an enemy is trying to manipulate and hack an AI system, the risks are even greater." (p. 140.)
- Serenko, Alexander (2010). "The development of an AI journal ranking based on the revealed preference approach" (http://www.aserenko.com/papers/JOI_Serenko_AI_Journal_Ranking_Published.pdf) (PDF). *Journal of Informetrics*. **4** (4): 447–59. doi:10.1016/j.joi.2010.04.001 (https://doi.org/10.1016%2Fj.joi.2010.04.001). Archived (https://web.archive.org/web/20131004215236/http://www.aserenko.com/papers/JOI_Serenko_AI_Journal_Ranking_Published.pdf) (PDF) from the original on 4 October 2013. Retrieved 24 August 2013.
- Serenko, Alexander; Michael Dohan (2011). "Comparing the expert survey and citation impact journal ranking methods: Example from the field of Artificial Intelligence" (http://www.aserenko.com/papers/JOI_AI_Journal_Ranking_Serenko.pdf) (PDF). *Journal of Informetrics*. **5** (4): 629–49. doi:10.1016/j.joi.2011.06.002 (https://doi.org/10.1016%2Fj.joi.2011.06.002). Archived (https://web.archive.org/web/20131004212839/http://www.aserenko.com/papers/JOI_AI_Journal_Ranking_Serenko.pdf) (PDF) from the original on 4 October 2013. Retrieved 12 September 2013.
- Tom Simonite (29 December 2014). "2014 in Computing: Breakthroughs in Artificial Intelligence" (http://www.technologyreview.com/news/533686/2014-in-computing-breakthroughs-in-artificial-intelligence/). *MIT Technology Review*.
- Sun, R. & Bookman, L. (eds.), *Computational Architectures: Integrating Neural and Symbolic Processes*. Kluwer Academic Publishers, Needham, MA. 1994.
- Taylor, Paul, "Insanely Complicated, Hopelessly Inadequate" (review of Brian Cantwell Smith, *The Promise of Artificial Intelligence: Reckoning and Judgment*, MIT, 2019, ISBN 978-0262043045, 157 pp.; Gary Marcus and Ernest Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust*, Ballantine, 2019, ISBN 978-1524748258, 304 pp.; Judea Pearl and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect*, Penguin, 2019, ISBN 978-0141982410, 418 pp.), *London Review of Books*, vol. 43, no. 2 (21 January 2021), pp. 37–39. Paul Taylor writes (p. 39): "Perhaps there is a limit to what a computer can do without knowing that it is manipulating imperfect representations of an external reality."
- Tooze, Adam, "Democracy and Its Discontents", *The New York Review of Books*, vol. LXVI, no. 10 (6 June 2019), pp. 52–53, 56–57. "Democracy has no clear answer for the mindless operation of bureaucratic and technological power. We may indeed be witnessing its extension in the form of artificial intelligence and robotics. Likewise, after decades of dire warning, the environmental problem remains fundamentally unaddressed.... Bureaucratic overreach and environmental catastrophe are precisely the kinds of slow-moving existential challenges that democracies deal with very badly.... Finally, there is the threat du jour: corporations and the technologies they promote." (pp. 56–57.)

External links

- "Artificial Intelligence" (http://www.iep.utm.edu/art-inte). *Internet Encyclopedia of Philosophy*.

- Thomason, Richmond. "Logic and Artificial Intelligence" (https://plato.stanford.edu/entries/logic-ai/). In Zalta, Edward N. (ed.). *Stanford Encyclopedia of Philosophy*.
- Artificial Intelligence (https://www.bbc.co.uk/programmes/p003k9fc), BBC Radio 4 discussion with John Agar, Alison Adam & Igor Aleksander (*In Our Time*, Dec. 8, 2005)

Sources

- This article incorporates text from a free content work. Licensed under C-BY-SA 3.0 IGO Text taken from *UNESCO Science Report: the Race Against Time for Smarter Development.* (https://unesdoc.unesco.org/ark:/48223/pf0000377433/PDF/377433eng.pdf.multi), Schneegans, S., T. Straza and J. Lewis (eds), UNESCO.